

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2005, 02:23:54 ; Search time 1671 Seconds  
(without alignments)  
666.948 Million cell updates/sec

Title: US-09-835-976B-115

Perfect score: 23

Sequence: 1 gtgtctaaggagaccaccagc 23

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hgt:\*  
3: gb\_in:\*  
4: gb\_on:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	23	100.0	63808	2 AC090554	AC090554 Homo sapi
C 2	23	100.0	150791	9 AL162458	AL162458 Human DNA
C 3	23	100.0	185931	2 AL133554	AL133554 Homo sapi
C 4	19.8	86.1	63808	2 AC090554	AC090554 Homo sapi
C 5	19.8	86.1	178050	2 AC142988	AC142988 Macaca mu
C 6	19	82.6	174956	10 AC131802	AC131802 Mus muscu
C 7	19	82.6	17519	2 AC145553	AC145553 Mus muscu
C 8	19	82.6	299377	2 AC127238	AC127238 Mus muscu
C 9	18.8	81.7	100711	9 AV623109	AV623109 Homo sapi
C 10	18.8	81.7	128625	10 AL837520	AL837520 Mouse DNA
C 11	18.8	81.7	149288	9 AC131444	AC131444 Homo sapi
C 12	18.8	81.7	162246	10 BX000428	BX000428 Mouse DNA
C 13	18.8	81.7	185212	2 BX572088	BX572088 Mus muscu
C 14	18.8	81.7	193572	10 AC140268	AC140268 Mus muscu
C 15	18.8	81.7	194490	10 AL845309	AL845309 Mouse DNA
C 16	18.8	81.7	200441	2 AC113895	AC113895 Rattus no
C 17	18.8	81.7	202992	9 AC068418	AC068418 Homo sapi
C 18	18.8	81.7	248196	2 AC117160	AC117160 Rattus no
C 19	18.8	81.7	277898	2 AC117083	AC117083 Rattus no

20	18.8	81.7	302912	2	AC116378	AC116378 Homo sapi
21	18.8	81.7	305000	9	HSXDPA	AL590762 Homo sapi
22	18.4	80.0	156917	2	AC150479	AC150479 Rhinolph
C 23	18.4	80.0	173766	2	AC129372	AC129372 Rattus no
C 24	18.4	80.0	202143	2	AC151542	AC151542 Callithri
C 25	18.4	80.0	209378	2	AC151545	AC151545 Callithri
26	18.4	80.0	238052	2	AC097824	AC097824 Rattus no
C 27	18.2	79.1	621	5	CR390750	CR390750 Gallus ga
C 28	18.2	79.1	15546	10	AL954858	AL954858 Mouse DNA
C 29	18.2	79.1	69515	2	AC099971	AC099971 Mus muscu
C 30	18.2	79.1	110000	2	AC115790_1	Continuation (2 of
C 31	18.2	79.1	139596	9	AC084837	AC084837 Homo sapi
C 32	18.2	79.1	140176	9	AC022188	AC022188 Homo sapi
C 33	18.2	79.1	143400	8	AC134927	AC134927 Oryza sat
C 34	18.2	79.1	144152	9	AL138954	AL138954 Human DNA
C 35	18.2	79.1	152580	9	AC024164	AC024164 Homo sapi
C 36	18.2	79.1	152779	9	AC018943	AC018943 Homo sapi
37	18.2	79.1	154010	2	AC042419	AC042419 Homo sapi
C 38	18.2	79.1	160517	10	AC132292	AC132292 Mus muscu
C 39	18.2	79.1	162506	2	AC115328	AC115328 Rattus no
C 40	18.2	79.1	168528	10	AC132255	AC132255 Mus muscu
C 41	18.2	79.1	168683	2	AC112923	AC112923 Mus muscu
C 42	18.2	79.1	171089	9	AC006368	AC006368 Homo sapi
C 43	18.2	79.1	171844	2	AC073430	AC073430 Homo sapi
44	18.2	79.1	175597	9	AC091195	AC091195 Homo sapi
C 45	18.2	79.1	177574	8	AC135424	AC135424 Oryza sat

#### ALIGNMENTS

RESULT 1	AC090554/c	63808 bp	DNA	linear	HTG 03-MAR-2001
LOCUS	Homo sapiens chromosome 17 clone RP11-191D16 map 17, LOW-PASS				
DEFINITION	SEQUENCE SAMPLING.				
ACCESSION	AC090554.1	GI:13194277			
VERSION	HTG; HTGS PHASE0.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 63808)				
AUTHORS	Birren, B., Linton, L., Nusbaum, C. and Lander, E.				
TITLE	Homo sapiens chromosome 17, clone RP11-191D16				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 63808)				
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Gaidan, J., Gagny, S., Glade, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuppback, R., Seaman, S., Severy, P., Sougnez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.				
TITLE	Direct Submission				
JOURNAL	Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
COMMENT	All repeats were identified using RepeatMasker:				

Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L10899  
 Center clone name: 191\_D\_16  
 -----

NOTE: This record contains 75 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

724:	Contig of 724 bp in length	21931	22030:	gap of 100 bp
725	824: gap of 100 bp	22031	22799:	contig of 769 bp in length
825	1636: contig of 812 bp in length	22800	22899:	gap of 100 bp
1637	1736: gap of 100 bp	22900	23649:	contig of 750 bp in length
1737	2433: contig of 697 bp in length	23650	23749:	gap of 100 bp
2434	2533: gap of 100 bp	23750	24473:	contig of 724 bp in length
2534	3256: contig of 723 bp in length	24474	24573:	gap of 100 bp
3257	3356: gap of 100 bp	24574	25377:	contig of 784 bp in length
3357	4122: contig of 766 bp in length	25378	25457:	gap of 100 bp
4123	4222: gap of 100 bp	25458	26221:	contig of 764 bp in length
4223	4966: contig of 744 bp in length	26222	26321:	gap of 100 bp
4967	5066: gap of 100 bp	26322	27091:	contig of 770 bp in length
5067	5818: contig of 752 bp in length	27092	27191:	gap of 100 bp
5819	5918: gap of 100 bp	27192	27927:	contig of 736 bp in length
5919	6623: contig of 705 bp in length	27928	28027:	gap of 100 bp
6624	6723: gap of 100 bp	28028	28782:	contig of 755 bp in length
6724	7535: contig of 812 bp in length	28783	28882:	gap of 100 bp
7536	7635: gap of 100 bp	28883	29601:	contig of 719 bp in length
7636	8413: contig of 778 bp in length	29602	29701:	gap of 100 bp
8414	9224: contig of 711 bp in length	30481:	contig of 780 bp in length	
9225	9324: gap of 100 bp	30482	30581:	gap of 100 bp
9325	10046: contig of 722 bp in length	30582	31361:	contig of 780 bp in length
10047	10146: gap of 100 bp	31362	31461:	gap of 100 bp
10147	10893: contig of 747 bp in length	31462	32147:	contig of 686 bp in length
10894	10993: gap of 100 bp	32148	32247:	gap of 100 bp
10994	11760: contig of 767 bp in length	32248	33025:	contig of 778 bp in length
11761	11860: gap of 100 bp	33026	33125:	gap of 100 bp
11861	12596: contig of 735 bp in length	33126	33839:	contig of 714 bp in length
12596	12695: gap of 100 bp	33840	33939:	gap of 100 bp
12696	13435: contig of 740 bp in length	33940	34680:	contig of 741 bp in length
13436	13535: gap of 100 bp	34681	34780:	gap of 100 bp
13536	14320: contig of 785 bp in length	34781	35528:	contig of 748 bp in length
14321	14420: gap of 100 bp	35529	35628:	gap of 100 bp
14421	15142: contig of 722 bp in length	35629	36401:	contig of 773 bp in length
15143	15242: gap of 100 bp	36402	36501:	gap of 100 bp
15243	16023: contig of 781 bp in length	36502	37266:	contig of 765 bp in length
16024	16123: gap of 100 bp	37267	37366:	gap of 100 bp
16124	16881: contig of 758 bp in length	37367	38120:	contig of 754 bp in length
16882	16981: gap of 100 bp	38121	38220:	gap of 100 bp
16982	17728: contig of 747 bp in length	38221	38952:	contig of 732 bp in length
17729	17828: gap of 100 bp	38953	39052:	gap of 100 bp
17829	18525: contig of 697 bp in length	39053	39812:	contig of 760 bp in length
18526	18625: gap of 100 bp	39813	39912:	gap of 100 bp
18626	19381: contig of 756 bp in length	39913	40649:	contig of 737 bp in length
19382	19481: gap of 100 bp	40650	40749:	gap of 100 bp
19482	20247: contig of 766 bp in length	40750	41524:	contig of 775 bp in length
20248	20347: gap of 100 bp	41525	41624:	gap of 100 bp
20348	21099: contig of 752 bp in length	41625	42384:	contig of 760 bp in length
21099	21199: gap of 100 bp	42385	42484:	gap of 100 bp
21199	21930: contig of 731 bp in length	42485	43241:	contig of 757 bp in length
21930		43242	43341:	gap of 100 bp
		43342	44105:	contig of 764 bp in length
		44106	44205:	gap of 100 bp
		44206	44977:	contig of 772 bp in length
		44978	45077:	gap of 100 bp
		45078	45855:	contig of 778 bp in length
		45856	45955:	gap of 100 bp
		45956	46731:	contig of 776 bp in length
		46732	46831:	gap of 100 bp
		46832	47504:	contig of 673 bp in length
		47505	47604:	gap of 100 bp
		47605	48379:	contig of 775 bp in length
		48380	48479:	gap of 100 bp
		48480	49230:	contig of 751 bp in length
		49231	49330:	gap of 100 bp
		49331	50130:	contig of 800 bp in length
		50131	50230:	gap of 100 bp
		50231	50989:	contig of 759 bp in length
		50990	51089:	gap of 100 bp
		51090	51832:	contig of 743 bp in length
		51833	51932:	gap of 100 bp
		51933	52684:	contig of 752 bp in length
		52685	52784:	gap of 100 bp

\* 52785 53585: contig of 801 bp in length  
 \* 53586 53685: gap of 100 bp  
 \* 53686 54453: contig of 768 bp in length  
 \* 54454 54553: gap of 100 bp  
 \* 54554 55274: contig of 721 bp in length  
 \* 55275 55374: gap of 100 bp  
 \* 55375 56145: contig of 771 bp in length  
 \* 56146 56245: gap of 100 bp  
 \* 56246 57010: contig of 765 bp in length  
 \* 57011 57110: gap of 100 bp  
 \* 57111 57835: contig of 725 bp in length  
 \* 57836 57935: gap of 100 bp  
 \* 57936 58674: contig of 739 bp in length  
 \* 58675 58774: gap of 100 bp

Query Match 100.0%; Score 23; DB 2; Length 63808;  
 Best Local Similarity 100.0%; Pred. No. 0.46; Mismatches 0; Indels 0; Gaps 0;  
 Matches 23; Conservative 0;

QY 1 GTGCTTAAAGGACACACG 23  
 |||||  
 Db 15433 GTGCTTAAAGGACACACG 15411

RESULT 2  
 AL162458/c  
 LOCUS  
 DEFINITION Human DNA sequence from clone RP11-465L10 on chromosome 20.  
 Contains 10 CpG islands, ESTs, STSs and GSSs. Contains the gene for a novel protein similar to Drosophila CG11399, the gene for a novel C2H2 type zinc finger protein similar to chicken FZF-1, a Ferritin light polypeptide (FTL) pseudogene, the MMP9 gene for matrix metalloproteinase 9 (gelatinase B, 92kD gelatinase, 92kD type IV collagenase) (CLG4B), a novel gene, the SLC12A5 gene for solute carrier family 12, (potassium-chloride transporter) member 5 (KIAA1176) and the 3' end of gene KIAA1637, complete sequence.  
 AL162458  
 AL162458.10 GI:9581599  
 HTG; collagenase; CpG island; ferritin; FTL; FZF-1; gelatinase; KCC2; KIAA1176; KIAA1637; matrix metalloproteinase; MMP9; SLC12A5; zinc finger.  
 SOURCE Homo sapiens (human)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 150791)  
 Williams, S.  
 Direct Submission  
 Submitted (26-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On Jul 28, 2000 this sequence version replaced gi:9367997.  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr20  
 IMPORTANT: This sequence is not the entire insert of clone RP11-465L10 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.  
 The true left end of clone RP4-599F21 is at 150692 in this sequence. The true right end of clone RP3-337O18 is at 100 in this sequence. This sequence was finished as follows unless otherwise

noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP11-465L10 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACE3.6.

FEATURES  
 source  
 1..150791  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="20"  
 /clone="RP11-465L10"  
 /clone\_lib="RPCI-11.2"  
 6..158  
 /note="AluX repeat: matches 1..153 of consensus"  
 161..469  
 /note="AluSg repeat: matches 1..308 of consensus"  
 475..761  
 /note="AluJb repeat: matches 12..294 of consensus"  
 985..1168  
 /note="MIR repeat: matches 5..202 of consensus"  
 1378..1685  
 /note="AluX repeat: matches 1..311 of consensus"  
 1723..1857  
 /note="AluJb repeat: matches 157..294 of consensus"  
 1858..2147  
 /note="AluJb repeat: matches 24..312 of consensus"  
 2148..2176  
 /note="AluJb repeat: matches 129..157 of consensus"  
 2177..2464  
 /note="AluSg repeat: matches 1..310 of consensus"  
 2465..2599  
 /note="AluJb repeat: matches 1..129 of consensus"  
 2760..2927  
 /note="MIR repeat: matches 94..261 of consensus"  
 2943..3188  
 /note="L2 repeat: matches 2183..2411 of consensus"  
 3189..3489  
 /note="AluX repeat: matches 2..304 of consensus"  
 3490..3735  
 /note="L2 repeat: matches 2411..2709 of consensus"  
 3856..4153  
 /note="AluSg repeat: matches 1..299 of consensus"  
 4420..4729  
 /note="AluJb repeat: matches 1..303 of consensus"  
 4954..5085  
 /note="MERS8A repeat: matches 85..219 of consensus"  
 5160..5318  
 /note="FRAM repeat: matches 1..159 of consensus"  
 5321..5429  
 /note="MIR repeat: matches 109..217 of consensus"  
 5457..5563  
 /note="FLAM\_C repeat: matches 1..107 of consensus"  
 5614..5904  
 /note="AluJb repeat: matches 1..288 of consensus"  
 5908..6052  
 /note="MIR repeat: matches 70..212 of consensus"  
 6080..6185  
 /note="MIR repeat: matches 89..197 of consensus"  
 7536..7693  
 /note="L2 repeat: matches 1817..1966 of consensus"  
 7694..7970  
 /note="AluJb repeat: matches 16..297 of consensus"  
 7971..8579  
 /note="L2 repeat: matches 1966..2742 of consensus"  
 8587..8885  
 /note="AluSg repeat: matches 1..299 of consensus"  
 9218..9503

```
/note="AluX repeat: matches 14. .296 of consensus"
9525. .9986
/note="L1MB8 repeat: matches 5624. .6108 of consensus"
9995. .10297
/note="AluSg repeat: matches 1. .309 of consensus"
10308. .10437
/note="AluJo repeat: matches 1. .137 of consensus"
10453. .10613
/note="AluSg repeat: matches 131. .304 of consensus"
10614. .10899
/note="AluSp repeat: matches 17. .304 of consensus"
complement(10640. .11142)
/note="match: GSS: Em:AQ512777"
complement(10653. .11145)
/note="match: GSS: Em:AQ114019"
10902. .10957
/note="L1ME1 repeat: matches 6108. .6163 of consensus"
11223. .11510
/note="AluSc repeat: matches 1. .290 of consensus"
12540. .12830
/note="AluX repeat: matches 1. .291 of consensus"
12835. .12935
/note="MER45 repeat: matches 22. .124 of consensus"
12948. .13243
/note="AluJo repeat: matches 3. .311 of consensus"
13539. .13839
/note="AluX repeat: matches 1. .300 of consensus"
13886. .14187
/note="L2 repeat: matches 2394. .2745 of consensus"
14188. .14472
/note="AluJo repeat: matches 2. .302 of consensus"
14473. .14573
/note="L2 repeat: matches 2294. .2394 of consensus"
14672. .14855
/note="L2 repeat: matches 2064. .2246 of consensus"
14865. .15007
/note="FLAM_C repeat: matches 1. .143 of consensus"
15008. .15300
/note="AluX repeat: matches 1. .296 of consensus"
15301. .15607
/note="AluX repeat: matches 1. .307 of consensus"
16056. .16359
/note="AluSg repeat: matches 1. .303 of consensus"
16373. .16662
/note="AluJb repeat: matches 1. .289 of consensus"
16666. .16933
/note="AluJo repeat: matches 6. .285 of consensus"
16950. .17249
/note="AluSc repeat: matches 1. .295 of consensus"
17261. .17291
/note="AluJo repeat: matches 113. .143 of consensus"
17292. .17577
/note="AluX repeat: matches 37. .298 of consensus"
17578. .17728
/note="AluJo repeat: matches 143. .291 of consensus"
17793. .18088
/note="AluX repeat: matches 1. .296 of consensus"
18210. .18514
/note="AluSp repeat: matches 1. .306 of consensus"
18625. .18757
/note="MIR repeat: matches 30. .153 of consensus"
18882. .19176
/note="AluSg repeat: matches 1. .292 of consensus"
19304. .19594
/note="AluX repeat: matches 1. .290 of consensus"
19852. .20068
/note="MIR repeat: matches 12. .245 of consensus"
complement(20480. .20897)
/note="match: GSS: Em:AQ616208"
20670. .21997
/note="CpG island"
evidence=not_experimental
21044. .34439
```

gene

## mrna

```
/gene="BA465L10.1"
Join(21044. .21220,23834. .24001,25397. .25539,25631. .25755,
26891. .27028,27225. .27355,27469. .27623,29513. .29660,
29756. .29839,30082. .30181,31284. .31446,32127. .32299,
32445. .32531,32616. .32800,33489. .33582,33679. .33854,
33940. .34439)
/gene="BA465L10.1"
```

Query Match 100.0%; Score 23; DB 9; Length 150791;

Best Local Similarity 100.0%; Pred. No. 0.43;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGCTTAAGGAGACACACAGC 23

Db 116416 GTGCTTAAGGAGACACACAGC 116394

## RESULT 3

```
AL133554
LOCUS Homo sapiens chromosome 20 clone RP1-91G13, 4 unordered pieces.
DEFINITION AL133554
ACCESSION AL133554
VERSION AL133554.28 GI:9930789
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
```

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Wall, M.

TITLE Direct Submission

JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Aug 27, 2000 this sequence version replaced gi:9863455.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: dj91G13

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 99% of reads

Chemistry: Dye-primer Big Dye; 0% of reads

Consensus quality: 184906 bases at least Q40

Consensus quality: 185263 bases at least Q30

Insert size: 185631; sum-of-contigs

Insert size: 151572; 42.7% error; agarose-fp

Quality coverage: 5.40x in Q20 bases; sum-of-contigs Quality

coverage: 6.65x in Q20 bases; agarose-fp

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 4 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence.

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 34065: contig of 34065 bp in length

\* 34066 34165: gap of 100 bp

\* 34166 79441: contig of 45276 bp in length

\* 79442 79541: gap of 100 bp

\* 79542 141187: contig of 61646 bp in length

\* 141188 141287: gap of 100 bp

\* 141288 185931: contig of 44644 bp in length.

Location/Qualifiers

1. .185931

/organism="Homo sapiens"

source

```

/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP1-91G13"
/clone_lib="RPC1-1"
1..34065
/note="assembly_fragment:02324
fragment_chain:1
clone_end:17
vector_side:left"
34166..79441
/note="assembly_fragment:01244
fragment_chain:1
clone_end:SP6
vector_side:right"
79542..141187
/note="assembly_fragment:01611"
141288..185931
/note="assembly_fragment:02960"

```

## ORIGIN

```

Query Match      100.0%; Score 23; DB 2; Length 185931;
Best Local Similarity 100.0%; Pred. No. 0.42; Mismatches 0; Indels 0; Gaps 0;
Matches 23; Conservative 0;

```

```

QY 1 GTGCTTAAGGAGACACACAGC 23
|||||
Db 107433 GTGCTTAAGGAGACACACAGC 107455

```

## RESULT 4

```

AC090554
LOCUS AC090554 63808 bp DNA linear HTG 03-MAR-2001
DEFINITION Homo sapiens chromosome 17 clone RP11-191D16 map 17, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC090554
VERSION AC090554.1 GI:13194277
KEYWORDS HTG; HTGS PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 63808)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Bouckhalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Laroque,K., Lamazares,R., Landers,T.,
Lehoczy,J., Levine,A., Lio,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylox,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schaefer,S., Schupbach,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Trigglio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

```

```

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: Li0899
Center clone name: 191_D_16

```

\* NOTE: This record contains 75 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

```

* 1 724: contig of 724 bp in length
* 725 824: gap of 100 bp
* 825 1636: contig of 812 bp in length
* 1637 1736: gap of 100 bp
* 1737 2433: contig of 697 bp in length
* 2434 2533: gap of 100 bp
* 2534 3256: contig of 723 bp in length
* 3257 3356: gap of 100 bp
* 3357 4122: contig of 766 bp in length
* 4123 4222: gap of 100 bp
* 4223 4966: contig of 744 bp in length
* 4967 5066: gap of 100 bp
* 5067 5818: contig of 752 bp in length
* 5819 5918: gap of 100 bp
* 5919 6623: contig of 705 bp in length
* 6624 6723: gap of 100 bp
* 6724 7535: contig of 812 bp in length
* 7536 7635: gap of 100 bp
* 7636 8413: contig of 778 bp in length
* 8414 8513: gap of 100 bp
* 8514 9224: contig of 711 bp in length
* 9225 9324: gap of 100 bp
* 9325 10046: contig of 722 bp in length
* 10047 10146: gap of 100 bp
* 10147 10893: contig of 747 bp in length
* 10894 10994: gap of 100 bp
* 10994 11760: contig of 767 bp in length
* 11761 11860: gap of 100 bp
* 11861 12595: contig of 735 bp in length
* 12596 13435: gap of 100 bp
* 13436 13535: contig of 740 bp in length
* 13536 14320: contig of 785 bp in length
* 14321 14420: gap of 100 bp
* 14421 15142: contig of 722 bp in length
* 15143 15242: gap of 100 bp
* 15243 16023: contig of 781 bp in length
* 16024 16123: gap of 100 bp
* 16124 16881: contig of 758 bp in length
* 16882 16981: gap of 100 bp
* 16982 17728: contig of 747 bp in length
* 17729 17828: gap of 100 bp
* 17829 18525: contig of 697 bp in length
* 18526 18626: gap of 100 bp
* 18626 19381: contig of 756 bp in length
* 19382 19481: contig of 100 bp
* 19482 20247: contig of 766 bp in length
* 20248 20347: gap of 100 bp
* 20348 21099: contig of 752 bp in length
* 21099 21100: gap of 100 bp
* 21100 21930: contig of 731 bp in length
* 21931 22030: gap of 100 bp

```

\* 22031 22799: contig of 769 bp in length  
\* 22800 22899: gap of 100 bp  
\* 22900 23649: contig of 750 bp in length  
\* 23650 23749: gap of 100 bp  
\* 23750 24473: contig of 724 bp in length  
\* 24474 24573: gap of 100 bp  
\* 24574 25357: contig of 784 bp in length  
\* 25358 25457: gap of 100 bp  
\* 25458 26221: contig of 764 bp in length  
\* 26222 26321: gap of 100 bp  
\* 26322 27031: contig of 770 bp in length  
\* 27032 27191: gap of 100 bp  
\* 27192 27927: contig of 736 bp in length  
\* 27928 28027: gap of 100 bp  
\* 28028 28782: contig of 755 bp in length  
\* 28783 28882: gap of 100 bp  
\* 28883 29601: contig of 719 bp in length  
\* 29602 30481: contig of 780 bp in length  
\* 30482 30581: gap of 100 bp  
\* 30582 31361: contig of 780 bp in length  
\* 31362 32141: gap of 100 bp  
\* 32142 32247: contig of 686 bp in length  
\* 32248 32247: gap of 100 bp  
\* 32249 33025: contig of 778 bp in length  
\* 33026 33125: gap of 100 bp  
\* 33126 33839: contig of 714 bp in length  
\* 33840 33939: gap of 100 bp  
\* 33940 34680: contig of 741 bp in length  
\* 34681 34780: gap of 100 bp  
\* 34781 35528: contig of 748 bp in length  
\* 35529 35628: gap of 100 bp  
\* 35629 36401: contig of 773 bp in length  
\* 36402 36501: gap of 100 bp  
\* 36502 37266: contig of 765 bp in length  
\* 37267 37366: gap of 100 bp  
\* 37368 38120: contig of 754 bp in length  
\* 38121 38220: gap of 100 bp  
\* 38221 38952: contig of 732 bp in length  
\* 38953 39052: gap of 100 bp  
\* 39053 39812: contig of 760 bp in length  
\* 39813 39912: gap of 100 bp  
\* 39913 40649: contig of 737 bp in length  
\* 40650 40749: gap of 100 bp  
\* 40750 41524: contig of 775 bp in length  
\* 41525 41624: gap of 100 bp  
\* 41625 42384: contig of 760 bp in length  
\* 42385 42484: gap of 100 bp  
\* 42485 43241: contig of 757 bp in length  
\* 43242 43341: gap of 100 bp  
\* 43342 44105: contig of 764 bp in length  
\* 44106 44205: gap of 100 bp  
\* 44206 44977: contig of 772 bp in length  
\* 44978 45077: gap of 100 bp  
\* 45078 45855: contig of 778 bp in length  
\* 45856 45955: gap of 100 bp  
\* 45956 46731: contig of 776 bp in length  
\* 46732 46831: gap of 100 bp  
\* 46832 47504: contig of 673 bp in length  
\* 47505 47604: gap of 100 bp  
\* 47605 48379: contig of 775 bp in length  
\* 48380 48479: gap of 100 bp  
\* 48480 49230: contig of 751 bp in length  
\* 49231 49330: gap of 100 bp  
\* 49331 50130: contig of 800 bp in length  
\* 50131 50230: gap of 100 bp  
\* 50231 50989: contig of 759 bp in length  
\* 50990 51089: gap of 100 bp  
\* 51090 51832: contig of 743 bp in length  
\* 51833 51932: gap of 100 bp  
\* 51933 52684: contig of 752 bp in length  
\* 52685 52784: gap of 100 bp  
\* 52785 53585: contig of 801 bp in length

\* 53586 53685: gap of 100 bp  
\* 53686 54453: contig of 768 bp in length  
\* 54454 54553: gap of 100 bp  
\* 54554 55274: contig of 721 bp in length  
\* 55275 55374: gap of 100 bp  
\* 55375 56145: contig of 771 bp in length  
\* 56146 56245: gap of 100 bp  
\* 56246 57010: contig of 765 bp in length  
\* 57011 57110: gap of 100 bp  
\* 57111 57835: contig of 725 bp in length  
\* 57836 57935: gap of 100 bp  
\* 57936 58674: contig of 739 bp in length  
\* 58675 58774: gap of 100 bp

Query Match 86.1%; Score 19.8; DB 2; Length 63808;  
Best Local Similarity 91.3%; Pred. No. 22;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTGTCTTAAGGAGACACACG 23  
Db 30634 GTGTCTTAAGGAGACACATATC 30656

RESULT 5  
AC142988  
LOCUS  
DEFINITION  
AC142988  
AC142988.1 GI:29567627  
HTG; HTGS PHASE2; HTGS PGI.  
VERSION  
Macaca mulatta (rhesus monkey)  
SOURCE  
Macaca mulatta  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
Cercopithecoidea; Macaca.  
1 (bases 1 to 178050)  
Csuros, M. and Milosavljevic, A.  
Pooled genomic indexing (PGI): mathematical analysis and experiment  
design  
(in) Guigo, R. and Gusfield, D. (Eds.);  
ALGORITHMS IN BIOINFORMATICS, SECOND INTERNATIONAL WORKSHOP, WABI  
2002, ROME, ITALY, SEPTEMBER 17-21, 2002, PROCEEDINGS: 10-28;  
Springer (2002)  
2 (bases 1 to 178050)  
Milosavljevic, A., Sodergren, E., Csuros, M., Li, B., Jackson, A.R.,  
Adams, C., Adio-Oduola, B., Ali-Isman, F.R., Allen, C., Alsbrooks, S.L.,  
Amaratunga, H.C., Are, J.R., Ayele, M., Banks, T., Barbaria, J.,  
Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J.,  
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,  
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,  
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,  
Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C., Cleveland, C.D.,  
Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,  
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,  
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,  
Dugan-Rocha, S., Durbin, K.J., Egan, A., Earnhart, C., Edwards, C.C.,  
Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D.,  
Elgaa, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J.,  
Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W.,  
Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K.,  
Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O.,  
Hodgson, A., Hogues, M., Hollway, C., Hollins, B., Honsi, F.,  
Howard, S., Huber, J., Hulyk, S., Hume, J., Ioshikhes, I., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
Karissom, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C.,  
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, J., Liu, W.,  
Loulsegh, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
Ma, J., Maheshwari, M., Mapua, P., Maravel, I., Martin, R.,  
Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P.,  
Meador, M., Mei, G., Merscher, S., Metzker, M., Miller, A., Miner, G.,  
Miner, Z., Mitchell, T., Mohabbat, K., Montgomery, K.T., Morgan, N.,  
Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N.,



REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 197519)  
Wilson.R.K.  
The sequence of Mus musculus clone  
Unpublished  
2 (bases 1 to 197519)  
Wilson.R.K.  
Direct Submission  
Submitted (19-JUL-2003) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
3 (bases 1 to 197519)  
Wilson.R.K.  
Direct Submission  
Submitted (30-APR-2004) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
On Apr 30, 2004 this sequence version replaced gi:40737111.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu  
Contact: submission@wustl.edu  
----- Project Information -----  
Center project name: M\_BA0106M01  
----- Summary Statistics -----  
Sequencing vector: M13; 0%  
Chemistry: Dye-terminator Big Dye; 100%  
Assembly program: Phrap; version 0.990319  
Consensus quality: 195074 bases at least Q40  
Consensus quality: 195598 bases at least Q30  
Consensus quality: 195993 bases at least Q20  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 6 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1505: contig of 1505 bp in length  
\* 1506 1605: gap of unknown length  
\* 1606 5809: contig of 4204 bp in length  
\* 5810 5909: gap of unknown length  
\* 5910 12827: contig of 6918 bp in length  
\* 12828 12927: gap of unknown length  
\* 12928 32221: contig of 19294 bp in length  
\* 32222 32321: gap of unknown length  
\* 32322 89970: contig of 57649 bp in length  
\* 89971 90071: gap of unknown length  
\* 90071 197519: contig of 107449 bp in length.

FEATURES  
source

1. .197519  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="7"  
/clone="RP23-106M1"  
1. .1505  
/note="assembly\_name:Contig17"  
1506. 5809  
/note="assembly\_name:Contig26"  
5910. 12827  
/note="assembly\_name:Contig27"  
12928. 32221  
/note="assembly\_name:Contig28"  
32322. 89970  
/note="assembly\_name:Contig29"  
90071. 197519  
/note="assembly\_name:Contig30"

## ORIGIN

Query Match 82.6%; Score 19; DB 2; Length 197519;  
Best Local Similarity 100.0%; Pred.No.52;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTGTCTTAAGGAGACACCA 19  
|||||  
Db 21310 GTGTCTTAAGGAGACACCA 21292  
|||||  
RESULT 8  
AC127238/c  
LOCUS  
DEFINITION  
AC127238  
VERSION  
AC127238.4 GI:52782653  
HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_ACTIVEFIN.  
KEYWORDS  
MUS musculus (house mouse)  
SOURCE  
MUS musculus  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Wilson.R.K.  
REFERENCE  
1 (bases 1 to 299377)  
TITLE  
The sequence of Mus musculus clone  
JOURNAL  
Unpublished  
2 (bases 1 to 299377)  
REFERENCE  
McPherson,J.D. and Waterston,R.H.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (14-JUL-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
JOURNAL  
3 (bases 1 to 299377)  
REFERENCE  
Wilson,R.K.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (28-SEP-2004) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
JOURNAL  
On Sep 28, 2004 this sequence version replaced gi:50284658.

## COMMENT

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu  
Contact: submission@wustl.edu  
----- Project Information -----  
Center project name: M\_BB0378P20  
----- Summary Statistics -----  
Sequencing vector: M13; 0%  
Chemistry: Dye-terminator Big Dye; 100%  
Assembly program: Phrap; version 0.990319  
Consensus quality: 293559 bases at least Q40  
Consensus quality: 295015 bases at least Q30  
Consensus quality: 296055 bases at least Q20  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 12 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1137: contig of 1137 bp in length  
\* 1138 1237: gap of unknown length  
\* 1238 2336: contig of 1099 bp in length  
\* 2337 2436: gap of unknown length  
\* 2437 3705: contig of 1269 bp in length  
\* 3706 3805: gap of unknown length  
\* 3806 11575: contig of 7770 bp in length  
\* 11576 11675: gap of unknown length



\* 11676 17379: contig of 5704 bp in length  
 \* 17380 17479: gap of unknown length  
 \* 26372: contig of 8893 bp in length  
 \* 26373 26472: gap of unknown length  
 \* 26473 45919: contig of 19447 bp in length  
 \* 45920 46019: gap of unknown length  
 \* 46020 72299: contig of 26280 bp in length  
 \* 72300 72399: gap of unknown length  
 \* 72400 104825: contig of 32426 bp in length  
 \* 104826 104925: gap of unknown length  
 \* 104926 174584: contig of 69659 bp in length  
 \* 174585 174684: gap of unknown length  
 \* 174685 237121: contig of 62437 bp in length  
 \* 237122 237222: gap of unknown length  
 \* 237222 299377: contig of 62156 bp in length.

## FEATURES

## source

1..299377

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosomes="7"

/clone="RP24-378P20"

1..1137

/notes="assembly\_name:Contig34"

1238..2336

/notes="assembly\_name:Contig37"

2437..3705

/notes="assembly\_name:Contig39"

3806..11575

/notes="assembly\_name:Contig45"

11676..17379

/notes="assembly\_name:Contig46"

17480..26372

/notes="assembly\_name:Contig47"

26473..45919

/notes="assembly\_name:Contig48"

46020..72299

/notes="assembly\_name:Contig49"

72400..104825

/notes="assembly\_name:Contig50"

clone\_end:77

vector\_side:left"

104926..174584

/notes="assembly\_name:Contig51"

clone\_end:SP6

vector\_side:right"

174685..237121

/notes="assembly\_name:Contig52"

237222..299377

/notes="assembly\_name:Contig53"

## ORIGIN

Query Match 82.8%; Score 19; DB 2; Length 299377;

Best Local Similarity 100.0%; Pred. No. 50;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGCTTTAAGGACACCA 19

Db 177456 GTGCTTTAAGGACACCA 177438

|||||

## RESULT 9

## LOCUS

AY623109 100711 bp DNA linear PRI 02-JUN-2004

## DEFINITION

Homo sapiens TAF1 RNA polymerase II, TATA box binding protein

(TBP)-associated factor, 250kDa (TAF1) gene, complete cds.

## ACCESSION

AY623109

## VERSION

AY623109.1 GI:47777654

## KEYWORDS

Homo sapiens (human)

## SOURCE

Homo sapiens

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 100711)

## AUTHORS

Livingston,R.J., Rieder,M.J., Chung,M.-W., Ritchie,T.K., Olson,A.N., Nguyen,C.P., Nguyen,D.A., Poel,C.L., Robertson,P.D., Schackwitz,W.S., Sherwood,J.K., Sherwood,J.K., Leithauser,B.J. and Nickerson,D.A.

## TITLE

Direct Submission

## JOURNAL

Submitted (12-MAY-2004) Genome Sciences, University of Washington,

1705 NE Pacific, Seattle, WA 98195, USA

## COMMENT

To cite this work please use: NIHES-SNPs, Environmental Genome Project, NIHES E515478, Department of Genome Sciences, Seattle, WA (URL: <http://egp.gs.washington.edu>).

## FEATURES

## source

1..100711

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

104..320

/rpt\_family="Alu"

/rpt\_type="dispersed

322..628

/rpt\_family="Alu"

/rpt\_type="dispersed

611

/frequency="0.01"

/replace="t"

639..770

/rpt\_family="Li"

/rpt\_type="dispersed

867..978

/rpt\_family="Alu"

/rpt\_type="dispersed

877

/frequency="0.07"

/replace="g"

1262..1506

/rpt\_family="Alu"

/rpt\_type="dispersed

1750

/frequency="0.01"

/replace="a"

1947

/frequency="0.07"

/replace="c"

2001..100192

/gene="TAF1"

Join(2001..2231,3239..3353,3794..3910,10907..11026,

12690..12931,13343..13561,13975..14193,14564..14771,

17483..17659,18276..18403,18501..18608,18731..18904,

19702..19875,20685..20789,23001..23201,23977..24118,

24478..24608,25325..25405,28310..28459,28616..28735,

29042..29217,29808..29986,32994..33207,34313..34478,

37269..37480,42379..42487,43315..43413,43715..43892,

57050..57117,58858..58980,59715..59807,59895..59979,

89911..89978,90482..90598,93982..94107,95293..95449,

96367..96544,99566..100192)

/gene="TAF1"

/product="TAF1 RNA polymerase II, TATA box binding protein

(TBP)-associated factor, 250kDa"

Join(2052..2231,3239..3353,3794..3910,10907..11026,

12690..12931,13343..13561,13975..14193,14564..14771,

17483..17659,18276..18403,18501..18608,18731..18904,

19702..19875,20685..20789,23001..23201,23977..24118,

24478..24608,25325..25405,28310..28459,28616..28735,

29042..29217,29808..29986,32994..33207,34313..34478,

37269..37480,42379..42487,43315..43413,43715..43892,

57050..57117,58858..58980,59715..59807,59895..59979,

89911..89978,90482..90598,93982..94107,95293..95449,

96367..96544,99566..99788)

/gene="TAF1"

/codon\_start=1

/product="TAF1 RNA polymerase II, TATA box binding protein

(TBP)-associated factor, 250kDa"

/protein\_id="AAT38105.1"

```
/db_xref="GI:47777655"
/translations="MGPGCDLLIRTAATITATAAIMSDTSDSDSAGGPGFSLAGFLFG
NINGAGLSESVLDDCKCKHLAGLGLSLITELTANEELTGTDGALVNDGQWVR
STEDAVDSIDINEADESRRVQOTMGLSLOPLCHSDYDEDDYDADCEDIDCKLMPPP
PPGPKWKDQDSITGIVENGEGIIPLSIIAPSSLASKEVDFSSSDSESEMGPOEA
TQAESDGLTLPGLMOHDATKLLPSVTELPFEPFGKGLRFLRFLFGFKGNVPSW
RSARRKKHRELIOEQIOEVCESEVSKLWNYDYAPPPPEQCLSDDEITM
MAPVESKFSQSTGDIQKVTDTKPRVAEWYGPALWYDMLGVPDGSDFYGFKLKRT
EHEPVIKSMIEBEKLENNGTOLLADENFLMTOLHWEDDIINDGEDVKKHGTQPO
RASLAGLPSSTRKAMAYNQGFPAATLDDDKPWYSIPIPNEDLVYRWEDNIWD
AQMPRLPPLVITDPDENLILEIPDEKERTSNPSKESKSSLKSLKRLIGKT
GVIKEPQQMSQPEVKDPWNLSDNYEYYPKQGLGTGFGNIIQHSIPAVLRQFF
PTHMGIKLQRPFLPKKYSFGALSGPQPHSVQPLPKHKKKAKMKEQERAGSGGE
MEFRTYQDLTGKDGDLILAEYSEENGLMMQVMATIKNYKRPKGPAGAPDCY
GETVYCHSPFLSGHPGLOAFENNLFRAPILYHMKPETFLLIIRTRQYVIRELV
RFDLONKVLSTSVLSTDDSSAEDSPFEMGKNIENMLONKKTSSOLSEREBQE
RKELOMLLAAGSAAGNNHRDDTASVTLSSSATGRCLKIYRPRDEGKEYVRCE
TVRKPAVIDAYVIRTKTBEFIRKFPALDEQREMRKERRIQOLRLKRNQBE
KLGPEPKPKKMRPDLKLCGACGAGHMTNKFCPLYYOTNAPPSPNPMATEEQ
EELEKTVIHNDEELIKVEGKIYLGKQILIESAEVRKSLVLPKQKLPVKKRR
VGTVHCDLNRPHKS IHRRTDPMVTLSSILESINDMRDLNTPYFHTPVNAKVVK
DYVYKITTDMDLQTLRENVKRLYPSREERFHELELVKNSATYNGPKHSLTQISQM
LQLCDEKLEKEDKLARLEKINPLDLDQVAFSPFILNIVTQKMAVPDSNPHHP
VKKFVPDIYKVINPMDLETIRKNIKSKYIQSRESFLODVLNLIANSVKYNGPSQY
TKTAQIBNVYQTLTEDEHLTQLEKICTAKAALAEALIESLDPMTPPGYTPQP
DLYDTVTSLSMRDASVFODESNMVLDIPSATPEKQVTOEGEDGDGLADBEETQV
QOASVLYEDLLAMEGDEBAGDEGNPFSALQESGSDSDVGSGGIRPKQPR
MLQENTRMDNEESNMSYEGDGEASHGLEDNLSYGSVEEPDKSNTQDTSFSSIG
GYEVSDEEBEERQSGFVLSQVHLSDEBDSDFHSIAGSDSLDSDSE"
2198
/variation
/genes="TAF1"
/frequency="0.01"
/replace="g"
2587
/variation
/genes="TAF1"
/frequency="0.01"
/replace="g"
2620
/variation
/genes="TAF1"
/frequency="0.03"
/replace="g"
2841
/variation
/genes="TAF1"
/frequency="0.01"
/replace="c"
3005
/variation
/genes="TAF1"
/frequency="0.93"
/replace="n"
3415
/variation
/rpt_family="Alu"
/rpt_type=dispersed
3433
/variation
/genes="TAF1"
/frequency="0.03"
/replace="t"
3557
/variation
/genes="TAF1"
/frequency="0.02"
/replace="t"
3572
/variation
/genes="TAF1"
/frequency="0.03"
/replace="t"
3654
/variation
/genes="TAF1"
/frequency="0.01"
/replace="a"
3677
/variation
/genes="TAF1"
/frequency="0.01"
/replace="t"
3994
/misc_feature
/genes="TAF1"
/note="Region not scanned for variation"
4330
/repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
4590
/repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
4890
/repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
5342
/repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
5684
/repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
6107
/repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
7410
/repeat_region
/rpt_family="MIR"
/rpt_type=dispersed
7890
/repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
8252
/repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
8505
/repeat_region
/rpt_family="MIR"
/rpt_type=dispersed
8607
/repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
8908
/repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
9682
/repeat_region
/variation
Query Match 81.7%; Score 18.8; DB 9; Length 100711;
Best Local Similarity 90.9%; Pred. No. 70;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 TGCTTTAGGAGACACACAGC 23
|||||
Db 87026 TGCTTTAGGAGACACATCAGC 87047
RESULT 10
AL837520/c
LOCUS 128625 bp DNA linear ROD 30-JUN-2003
DEFINITION Mouse DNA sequence from clone Rp23-41204 on chromosome 2, complete
sequence.
ACCESSION AL837520
VERSION AL837520.26 GI:32398638
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 128625)
AUTHORS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Bates, K.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Jul 1, 2003 this sequence version replaced gi:31335607.
```

Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Information on the WormPeP database can be found at

[http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) RP23-41204 is

from the RP23-41204 Mouse BAC Library

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6.

#### FEATURES

source

1. 128625  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosomes="2"  
/clone="RP23-41204"  
/clone\_lib="RP23-23"

#### ORIGIN

Query Match 81.7%; Score 18.8; DB 10; Length 128625;  
Best Local Similarity 90.9%; Pred. No. 69;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTGCTTAAGGACACACACAG 22

Db 65566 GTGCTTCTGGAGACACACAG 65545

#### RESULT 11

AC113144  
LOCUS AC113144  
DEFINITION Homo sapiens chromosome 17, clone RP11-45G12, complete sequence.  
AC113144  
VERSION AC113144.11 GI:24961989

KEYWORDS HTG.

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 149288)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,

Brown, A., Camarata, J., Chang, J., Chazaro, B.,

Choepe, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,

Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,

Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,

Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,

Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,

Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (14-NOV-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Nov 14, 2002 this sequence version replaced gi:24462409.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,  
Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C.,  
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,  
McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,  
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
Rosa, S., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S.,  
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, K., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

#### Direct Submission

Submitted (25-FEB-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 149288)

#### JOURNAL

#### REFERENCE

#### AUTHORS

Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,  
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,  
Camarata, J., Chang, J., Chazaro, B., Choepe, Y., Collymore, A.,  
Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,  
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,  
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J.,  
Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T.,  
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,  
Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R.,  
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,  
Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,  
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

#### Direct Submission

Submitted (01-NOV-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 149288)

#### JOURNAL

#### REFERENCE

#### AUTHORS

Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,  
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,  
Camarata, J., Chang, J., Chazaro, B., Choepe, Y., Collymore, A.,  
Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,  
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,  
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J.,  
Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T.,  
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,  
Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R.,  
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,  
Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,  
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

#### Direct Submission

Submitted (14-NOV-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Nov 14, 2002 this sequence version replaced gi:24462409.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information  
 Center project name: L26065  
 Center clone name: 45\_G\_12  
 -----

## FEATURES

## Location/Qualifiers

```

source
1. .149288
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-45G12"
/clone_lib="RPCI-11 Human Male BAC"
1181. .1196
/notes="<30 qual SINGL region"
1219. .1225
/notes="<30 qual SINGL region"
1232. .1241
/notes="<30 qual SINGL region"
1390. .1397
/notes="<30 qual SINGL region"
1853. .2299
/notes="single clone coverage"
repeat_region
/rpt_family="MLT1D"
2405. .2848
/rpt_family="MSTD"
repeat_region
/rpt_family="LTR49"
complement(2901. .2979)
repeat_region
complement(3065. .3231)
/rpt_family="LTR49-int"
repeat_region
complement(3230. .5505)
/rpt_family="LTR49-int"
repeat_region
complement(5628. .5784)
/rpt_family="AluJo"
repeat_region
complement(5785. .6086)
/rpt_family="AluSp"
repeat_region
complement(6087. .6235)
/rpt_family="AluJo"
6277. .6395
/rpt_family="AluSx"
6396. .6448
/rpt_family="AluSx"
/rpt_family="(TAAA)n"
6449. .6609
/rpt_family="AluSx"
repeat_region
complement(6611. .6881)
/rpt_family="AluJb"
6664
/notes="probably A"
unsure
complement(6866. .6870)
/notes="<30 qual SINGL region"
repeat_region
complement(6901. .7063)
/rpt_family="LTR49-int"
repeat_region
complement(7280. .7364)
/rpt_family="LTR49-int"
7365. .7659
/rpt_family="AluSc"
repeat_region
complement(7660. .8154)
/rpt_family="LTR49-int"
repeat_region
complement(8256. .8589)
/rpt_family="LPA9"
repeat_region
complement(8590. .8992)
/rpt_family="LTR49-int"
repeat_region
complement(8993. .9290)
/rpt_family="AluSx"
repeat_region
complement(9291. .9324)
/rpt_family="LTR49-int"
repeat_region
complement(9422. .9759)
/rpt_family="LTR49-int"
repeat_region
complement(9762. .10309)
/rpt_family="LTR49"
repeat_region
complement(10353. .10478)
/rpt_family="HAL1"
10479. .11077
/rpt_family="AluJb"

```

```

repeat_region complement(10778. .10848)
/rpt_family="HAL1"
repeat_region complement(10850. .10887)
/rpt_family="LIMC4"
10888. .11179
/rpt_family="AluSx"
repeat_region complement(11180. .11904)
/rpt_family="LIMC4"
repeat_region complement(11964. .12258)
/rpt_family="LIMED"
repeat_region complement(12541. .12807)
/rpt_family="LIMA9"
12808. .12841
/rpt_family="(TTTA)n"
repeat_region complement(12842. .13125)
/rpt_family="AluSx"
repeat_region complement(13126. .13239)
/rpt_family="LIMA9"
repeat_region complement(13269. .13701)
/rpt_family="Tigger2"
repeat_region complement(13700. .13768)
/rpt_family="Tigger2"
repeat_region complement(13769. .13936)
/rpt_family="LIMA9"
13967. .14140
/rpt_family="LTR48"
repeat_region 14156. .14469
/rpt_family="LTR48B"
repeat_region complement(14470. .14689)
/rpt_family="LIMA9"
repeat_region complement(15000. .15229)
/rpt_family="LIM4"
repeat_region complement(15230. .15542)
/rpt_family="AluSx"
repeat_region complement(15543. .15595)

```

Query Match 81.7%; Score 18.8; DB 9; Length 149288;  
 Best Local Similarity 90.9%; Pred. No. 68;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTGTCTTAAGGAGACACACAG 22  
 ||||| ||||| ||||| ||||| |||||  
 Db 97199 GTGTCTTAGGAGACACACAG 97220

## RESULT 12

```

BX000428 BX000428 162246 bp DNA linear ROD 09-AUG-2003
LOCUS Mouse DNA sequence from clone RP23-474G7 on chromosome X, complete
DEFINITION sequence.
ACCESSION BX000428
VERSION BX000428.10 GI:33569160
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 162246)
Chapman,J.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (09-AUG-2003) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT
On Aug 9, 2003 this sequence version replaced gi:33412166.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

```

Query Match	81.7%;	Score 18.8;	DB 2;	Length 185212;
Best Local Similarity	90.9%;	Pred. No. 67;		
Matches	20;	Conservative	0;	Mismatches 2;
				Indels 0;
				Gaps 0;
QY	2	TGTCCTTAAGGACACACACAGC	23	
Db	78682	TGTCCTTAAGGACACACACACC	78661	
RESULT 14				
AC140268				
LOCUS				
DEFINITION	AC140268	193572 bp	DNA	linear
	Mus musculus BAC clone RP23-321K4	from chromosome 7,	complete	
ACCESSION	AC140268			
VERSION	AC140268.2	GI:33086434		
KEYWORDS	HTG.			
SOURCE	Mus musculus	(house mouse)		
ORGANISM	Mus musculus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1	(bases 1 to 193572)		
AUTHORS	Abbott,S.			
TITLE	The sequence of Mus musculus BAC clone RP23-321K4			
JOURNAL	Unpublished (2001)			
REFERENCE	2	(bases 1 to 193572)		
AUTHORS	Wilson,R.			
TITLE	Sequencing of Mus musculus			
JOURNAL	Unpublished (2001)			
REFERENCE	3	(bases 1 to 193572)		
AUTHORS	McPherson,J.D. and Waterston,R.H.			

TITLE Direct Submission  
 JOURNAL Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 REFERENCE 4 (bases 1 to 193572)  
 AUTHORS Wilson,R.K.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-JUL-2003) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 REFERENCE 5 (bases 1 to 193572)  
 AUTHORS Wilson,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-NOV-2003) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT On Jul 21, 2003 this sequence version replaced gi:28475512.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu>  
 Contact: [submissions@wustl.edu](mailto:submissions@wustl.edu)  
 ----- Summary Statistics  
 Center project name: M\_BA0321K04  
 -----

NOTICE: This sequence may not represent the entire insert of this  
 clone. It may be shorter because we only sequence overlapping  
 clone sections once, or longer because we provide a small overlap  
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate  
 chemistry, or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by sequence  
 from more than one subclone; and the assembly was confirmed by  
 restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren,  
 Department of Genetics, Washington University, St. Louis MO. For  
 additional information about the map position of this sequence, see  
<http://genome.wustl.edu>

#### SOURCE INFORMATION:

The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa  
 and Minako Tatenio in the laboratory of Pieter de Jong  
 (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or  
 brain genomic DNA. The clone and detailed information can be  
 obtained from Research Genetics, Inc. (<http://www.resgen.com>) or  
 Pieter de Jong and coworkers at <http://www.chori.org>

#### NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

#### FEATURES

source	Location/Qualifiers
1. .193572	
/organism="Mus musculus"	
/mol_type="genomic DNA"	
/db_xref="taxon:10090"	
/chromosome="7"	
/map="7"	
/clone="RP23-321K4"	
/clone_lib="RPCI-23"	
54. .200	
/rpt_family="Alu"	
572. .628	
/rpt_family="Alu"	
657. .776	
/rpt_family="Alu"	
834. .1023	
/rpt_family="B2"	
1064. .1137	
/rpt_family="Alu"	
1138. .1164	
/rpt_family="B4"	
repeat_region	1294. .1574
	/rpt_family="B4"
repeat_region	2448. .2622
	/rpt_family="B2"
repeat_region	2625. .2754
	/rpt_family="Alu"
repeat_region	2762. .2947
	/rpt_family="B2"
repeat_region	3633. .3735
	/rpt_family="Alu"
repeat_region	3736. .3811
	/rpt_family="B4"
repeat_region	4063. .4155
	/rpt_family="ID"
repeat_region	4190. .4268
	/rpt_family="Alu"
repeat_region	4293. .4410
	/rpt_family="L1"
repeat_region	4411. .4607
	/rpt_family="B2"
repeat_region	4611. .4756
	/rpt_family="Alu"
repeat_region	4802. .4950
	/rpt_family="Alu"
repeat_region	5022. .5163
	/rpt_family="Alu"
repeat_region	5235. .5343
	/rpt_family="Alu"
repeat_region	5344. .5370
	/rpt_family="B4"
repeat_region	5562. .5679
	/rpt_family="B4"
repeat_region	5689. .6308
	/rpt_family="RMR1B"
repeat_region	6297. .6427
	/rpt_family="B2"
repeat_region	6620. .6813
	/rpt_family="MaLR"
repeat_region	7108. .7277
	/rpt_family="MaLR"
repeat_region	7348. .7463
	/rpt_family="ERVK"
repeat_region	7529. .7734
	/rpt_family="MaLR"
repeat_region	8230. .8441
	/rpt_family="MaLR"
repeat_region	8444. .8775
	/rpt_family="MaLR"
repeat_region	10057. .10145
	/rpt_family="MIR"
repeat_region	10467. .10538
	/rpt_family="B2"
repeat_region	10552. .10558
	/rpt_family="B4"
repeat_region	12887. .13023
	/rpt_family="Alu"
repeat_region	13180. .13290
	/rpt_family="B4"
repeat_region	13278. .13446
	/rpt_family="B2"
repeat_region	13453. .13625
	/rpt_family="B4"
repeat_region	13691. .13873
	/rpt_family="B2"
repeat_region	13997. .14148
	/rpt_family="Alu"
repeat_region	14393. .14553
	/rpt_family="B4"
repeat_region	14594. .14759
	/rpt_family="B2"
trna	complement(14681. .14753)
	/product="tRNA-Ser"
	/note="Likely pseudogene (HMM Sc=37.06 / Sec struct

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-376020 is from the RPCI-23 Mouse PAC Library  
 constructed by the group of Pieter de Jong.  
 For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

# FEATURES

Location/Qualifiers  
 1..194490  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="2"  
 /clone="RP23-376020"  
 /clone\_lib="RPCI-23"

# ORIGIN

Query Match 81.7%; Score 18.8; DB 10; Length 194490;  
 Best Local Similarity 90.9%; Pred. No. 67;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 TGTCTTTAAGGAGACACACAGC 23  
 |||||  
 Db 101086 TGCCTTAGGAGACACACAGC 101107

Search completed: March 9, 2005, 07:53:30  
 Job time: 1677 secs

Sc=11.50"  
 15947..16096  
 /rpt\_family="Alu"  
 16540..16752  
 /rpt\_family="B4"  
 16765..16911  
 /rpt\_family="Alu"  
 18092..18293  
 /rpt\_family="B2"  
 18308..18449  
 /rpt\_family="Alu"  
 18587..18746  
 /rpt\_family="MalR"  
 18641..18776  
 /rpt\_family="B4"  
 18818..19040  
 /rpt\_family="MER2\_type"  
 19630..19835  
 /rpt\_family="B2"  
 21195..21874  
 /rpt\_family="RMER12"  
 21942..22091  
 /rpt\_family="Alu"  
 23102..23231  
 /rpt\_family="L1"  
 23288..23424  
 /rpt\_family="B4"  
 23610..23789  
 /rpt\_family="L1"  
 23794..23904  
 /rpt\_family="Alu"  
 24726..24876  
 /rpt\_family="B2"  
 25081..25226

Query Match 81.7%; Score 18.8; DB 10; Length 193572;  
 Best Local Similarity 90.9%; Pred. No. 67;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GTGTCTTAAGGAGACACACAGC 22  
 |||||  
 Db 87154 GTGTCTAAGGAGACACACAGC 87175

RESULT 15  
 AL845309 194490 bp DNA linear ROD 22-AUG-2002  
 LOCUS Mouse DNA sequence from clone RP23-376020 on chromosome 2, complete  
 DEFINITION sequence.  
 ACCESSION AL845309  
 VERSION AL845309.2 GI:22474428  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 194490)  
 Williams, S.  
 Direct Submission  
 Submitted (22-AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 On Aug 23, 2002 this sequence version replaced gi:22205004.  
 COMMENT  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: humquery@sanger.ac.uk  
 -----

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2005, 02:14:38 ; Search time 427 Seconds  
(without alignments)  
318.862 Million cell updates/sec

Title: US-09-835-976B-115

Perfect score: 23  
Sequence: 1 gtgtcttaaggagacaccacagc 23

Scoring table: IDENTITY\_NUC

Gapop 10\_0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*

1: Geneseqn1980s:\*

2: Geneseqn1990s:\*

3: Geneseqn2000s:\*

4: Geneseqn2001as:\*

5: Geneseqn2001bs:\*

6: Geneseqn2002as:\*

7: Geneseqn2002bs:\*

8: Geneseqn2003as:\*

9: Geneseqn2003bs:\*

10: Geneseqn2003cs:\*

11: Geneseqn2003ds:\*

12: Geneseqn2004as:\*

13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	100.0	23	4	AAS59492 Human KCC
2	18.2	79.1	231004	12	ADQ97855 Mouse can
3	18	78.3	487	6	ABK71639 Human dit
4	18	78.3	1287	12	ADQ00741 Novel hum
5	18	78.3	1287	12	ADN99172 Novel hum
6	18	78.3	1726	5	AAS72172 DNA encod
7	18	78.3	1766	12	ACH87437 Human gen
8	18	78.3	2077	3	Aaz52430 HTRM clon
9	17.8	77.4	1646	8	ABX15830 cDNA enco
10	17.8	77.4	1646	12	ADJ18998 Soybean s
11	17.8	77.4	52302	9	ADA02738 Human CN
12	17.8	77.4	52302	10	ADB72476 Human CN
13	17.8	77.4	52302	10	ADC85218 Human Ccn
14	17.8	77.4	52302	12	ADM74333 Human car
15	17.8	77.4	144035	11	ACN44062 Human gen
16	17.2	74.8	146	12	ADK43275 Murine pr
17	17.2	74.8	761	10	ACD94539 Human col
18	17.2	74.8	1872	3	AAA47169 DNA seque
19	17.2	74.8	2075	2	AAV16467 Nucleotid
20	17.2	74.8	2091	2	AAI18839 Streptoco

21	17.2	74.8	2872	2	AAQ21001 Murine re
22	17.2	74.8	2872	2	AAT36881 Murine re
23	17.2	74.8	2872	2	AAX29853 Mouse rec
24	17.2	74.8	2872	12	ADJ92679 Murine pr
25	17.2	74.8	3070	12	ADK43187 Murine pr
26	17.2	74.8	74849	11	ACN44932_3 Continuation (4 of
27	17.2	74.8	84073	11	ACN44322 Human gen
28	17.2	74.8	202251	11	ACN44504 Mouse gen
29	16.8	73.0	417	5	ABV17885 Human pro
30	16.8	73.0	424	10	ADD49392 Human lun
31	16.8	73.0	534	5	ABV47675 Human pro
32	16.8	73.0	577	9	ACH39337 Human foe
33	16.8	73.0	1203	12	ADQ87507 Human tum
34	16.8	73.0	1203	12	ADQ84889 Human tum
35	16.8	73.0	1203	13	ADQ87165 Human tum
36	16.8	73.0	3135	10	ADE60162 Human gen
37	16.8	73.0	13058	4	AAK79325 Human imm
38	16.8	73.0	13058	5	AAS29194 Genomic s
39	16.8	73.0	13058	6	ABS68334 Human DNA
40	16.8	73.0	13058	10	ADC25456 Human cdn
41	16.8	73.0	25229	12	ADQ97226 Mouse can
42	16.8	73.0	29111	11	ACN44424 Mouse gen
43	16.8	73.0	107818	9	AAL57572 Human pho
44	16.8	73.0	340449	8	AAL52198 Human sec
45	16.6	72.2	512	6	ABK62773 Rat seque

ALIGNMENTS

RESULT 1  
AAS59492  
ID AAS59492 standard; DNA; 23 BP.

AC AAS59492;

DT 16-JAN-2002 (first entry)

DE Human KCC2 polymorphism detection primer #2.

KW Human; ss; potassium-chloride cotransporter; epilepsy; hypertension; KCC;  
KW hypotensive; anticonvulsant; antianaemic; sickle-cell anaemia;  
KW gene therapy; PCR primer.

OS Homo sapiens.

PN WO200179525-A2.

XX 25-OCT-2001.

PF 16-APR-2001; 2001WO-US012395.

PR 14-APR-2000; 2000US-0197350P.

XX (UYVA-) UNIV VANDERBILT.  
(GAMB/) GAMBA G.

XX Gamba G, Mount DB, Delpire E, George AL;

DR WPI; 2001-611726/70.

XX New isolated and purified potassium-chloride cotransporter polypeptide  
PT for detecting a modulator of potassium-chloride cotransport that can  
PT treat epilepsy, sickle cell anemia, and hypertension.

XX Example 26; Page 347; 352pp; English.

XX The invention relates to an isolated and purified, biologically active  
CC KCC2, KCC3, or KCC4 potassium-chloride cotransporter polypeptide and  
CC polynucleotides encoding them. KCC and the nucleic acid encoding KCC are  
CC used to produce an antibody against KCC, which can be used to detect KCC.  
CC The antibody and nucleic acid are used in kits to detect KCC and the  
CC nucleic acid encoding KCC. KCC and nucleic acid encoding it are used to

CC screen for candidate substances that can modulate the activity of KCC.  
 CC KCC and the nucleic acid encoding it can be used to modulate potassium-  
 CC chloride cotransport in a vertebrate. Therapeutic compositions comprising  
 CC modulators of the activity of KCC (preferably antibodies or antisense  
 CC oligonucleotides) are used to treat epilepsy, sickle cell anaemia, and  
 CC hypertension (also via gene therapy using the KCC polynucleotide). The  
 CC present sequence is a PCR primer used to detect polymorphisms in a DNA  
 CC sequence encoding a KCC of the invention  
 XX  
 SQ Sequence 23 BP; 7 A; 6 C; 6 G; 4 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 23; DB 4; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 0.36;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTGCTTTAAGGAGACACACAGC 23  
 |||||  
 Db 1 GTGCTTTAAGGAGACACACAGC 23  
 |||||  
 RESULT 2  
 ADQ97855  
 ID ADQ97855 standard; DNA; 231004 BP.  
 AC  
 AC ADQ97855;  
 XX  
 XX 07-OCT-2004 (first entry)  
 XX  
 DE Mouse cancer associated sequence MD11-013, SEQ ID 832.  
 XX  
 KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.  
 XX  
 XX Mus musculus.  
 XX  
 PN WO2004060304-A2.  
 XX  
 XX 22-JUL-2004.  
 XX  
 XX 22-DEC-2003; 2003WO-US041389.  
 XX  
 XX 27-DEC-2002; 2002US-00330773.  
 XX  
 XX (SAGR-) SAGRES DISCOVERY INC.  
 XX  
 XX Morris DW, Malandro MS;  
 XX  
 DR WPI; 2004-543781/52.  
 XX  
 XX New isolated cancer associated nucleic acids comprising at least 10  
 XX contiguous nucleotides, useful for diagnosing, preventing and/or treating  
 XX cancers such as leukemia and lymphoma.  
 XX  
 PS Claim 1; SEQ ID NO 832; 199pp; English.  
 XX  
 XX The present invention relates to cancer associated sequences (ADQ97025-  
 CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or  
 CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 231004 BP; 66840 A; 40969 C; 41531 G; 69116 T; 0 U; 12548 Other;  
 Query Match 79.1%; Score 18.2; DB 12; Length 231004;  
 Best Local Similarity 87.0%; Pred. No. 2.2e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 GTGCTTTAAGGAGACACACAGC 23  
 |||||  
 Db 178598 GTGCTTGAAGGAACACACATC 178620  
 |||||

RESULT 3

ABK71639  
 ID ABK71639 standard; cDNA; 487 BP.  
 XX  
 AC ABK71639;  
 XX  
 DT 30-JUL-2002 (first entry)  
 XX  
 XX Human dithp polynucleotide #105.  
 XX  
 XX Human; dithp; diagnostic and therapeutic polynucleotide; gene; ss; bone;  
 KW cell proliferative disorder; cancer; tumour; autoimmune disorder; brain;  
 KW inflammatory disorder; viral infection; bacterial infection; seizure;  
 KW fungal infection; parasitic infections; developmental disorder; breast;  
 KW endocrine disorder; metabolic disorder; neurological disorder; cervix;  
 KW gastrointestinal disorder; transport disorder; gene therapy; kidney;  
 KW adrenal gland; bone marrow; lung; ovary; pancreas; prostate; spleen;  
 KW skin; testis; thymus.  
 XX  
 XX Homo sapiens.  
 XX  
 PN WO200220754-A2.  
 XX  
 PD 14-MAR-2002.  
 XX  
 PF 29-AUG-2001; 2001WO-US027127.  
 XX  
 XX 05-SEP-2000; 2000US-0229747P.  
 PR 05-SEP-2000; 2000US-0229748P.  
 PR 05-SEP-2000; 2000US-0229749P.  
 PR 05-SEP-2000; 2000US-0229750P.  
 PR 05-SEP-2000; 2000US-0229751P.  
 PR 05-SEP-2000; 2000US-0230583P.  
 PR 06-SEP-2000; 2000US-0230505P.  
 PR 06-SEP-2000; 2000US-0230514P.  
 PR 06-SEP-2000; 2000US-0230515P.  
 PR 06-SEP-2000; 2000US-0230517P.  
 PR 06-SEP-2000; 2000US-0230518P.  
 PR 06-SEP-2000; 2000US-0230519P.  
 PR 06-SEP-2000; 2000US-0230595P.  
 PR 06-SEP-2000; 2000US-0230597P.  
 PR 06-SEP-2000; 2000US-0230598P.  
 PR 06-SEP-2000; 2000US-0230610P.  
 PR 06-SEP-2000; 2000US-0230865P.  
 PR 06-SEP-2000; 2000US-0230988P.  
 PR 07-SEP-2000; 2000US-0230951P.  
 PR 07-SEP-2000; 2000US-0231163P.  
 PR 07-SEP-2000; 2000US-0231167P.  
 XX  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX  
 XX Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;  
 PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;  
 PI Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;  
 PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;  
 PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;  
 XX  
 XX WPI; 2002-383054/41.  
 DR P-PSDB; ABG60048.  
 XX  
 XX An isolated polynucleotide useful in diagnostics and therapeutics.  
 PT  
 PS Claim 1; Page 454; 686pp; English.  
 XX  
 XX The invention relates to human diagnostic and therapeutic (dithp)  
 CC polynucleotides and their associated polypeptides (DITHP polypeptides).  
 CC The sequences of the invention are used in the treatment and diagnosis of  
 CC cell proliferative disorders (e.g. atherosclerosis, cirrhosis), cancers  
 CC (e.g. tumours of the adrenal gland, bone, bone marrow, brain, breast,  
 CC cervix, kidney, lung, ovary, pancreas, prostate, skin, spleen, testis or  
 CC thymus), autoimmune/inflammatory disorders (e.g. asthma, bronchitis,  
 CC psoriasis, osteoporosis), viral infections, bacterial infections, fungal  
 CC infections, parasitic infections, developmental disorders (e.g. anaemia,

CC epilepsyl, seizure disorders (e.g. cerebral palsy, spina bifida),  
 CC endocrine disorders (e.g. thrombosis, aneurysm), metabolic disorders  
 CC (e.g. obesity, diabetes), neurological disorders (e.g. stroke,  
 CC amyotrophic lateral sclerosis, multiple sclerosis), gastrointestinal  
 CC disorders (e.g. ulcerative colitis, lysinuria) and transport disorders  
 CC (e.g. myotonic dystrophy, catatonla, peripheral neuropathy). Sequences  
 CC ABK71535-ABK71809 represent human dithp polynucleotides of the invention  
 XX

Sequence 487 BP; 162 A; 94 C; 101 G; 130 T; 0 U; 0 Other;

Query Match 78.3%; Score 18; DB 6; Length 487;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TCTTAAGGAGACACCACA 21

|||||

Db 82 TCTTAAGGAGACACCACA 99

#### RESULT 4

ADO00741

ID ADO00741 standard; cDNA; 1287 BP.

XX AC

ADO00741;

XX AC

DT 29-JUL-2004 (first entry)

XX XX

DE Novel human cDNA sequence #1556.

XX KW

ds; gene; anti-inflammatory; dermatological; neuroprotective;

immunomodulator; antibacterial; virucide; antiproliferative; cytostatic;

KW gene therapy; vaccine; inflammatory; CNS; immune disorder; cancer;

KW psoriasis; diabetes; early aging; hormonal imbalance;

KW ischemic heart disease; ulcerative colitis.

XX XX

OS Homo sapiens.

XX XX

PN WO2004038003-A2.

XX XX

PD 06-MAY-2004.

XX XX

PF 24-OCT-2003; 2003WO-US033947.

XX XX

PR 25-OCT-2002; 2002US-0421061P.

XX PR

PR 25-OCT-2002; 2002US-0421080P.

XX PR

PR 25-OCT-2002; 2002US-0421552P.

XX PR

PR 25-OCT-2002; 2002US-0421614P.

XX PR

PR 30-OCT-2002; 2002US-0422177P.

XX PR

PR 30-OCT-2002; 2002US-0422178P.

XX PR

PR 15-NOV-2002; 2002US-0426355P.

XX PR

PR 15-NOV-2002; 2002US-0426384P.

XX PR

PR 15-NOV-2002; 2002US-0426394P.

XX PR

PR 15-NOV-2002; 2002US-0426430P.

XX PR

PR 15-NOV-2002; 2002US-0426916P.

XX PR

PR 27-NOV-2002; 2002US-0429224P.

XX PR

PR 27-NOV-2002; 2002US-0429275P.

XX PR

PR 27-NOV-2002; 2002US-0429302P.

XX PR

PR 27-NOV-2002; 2002US-0429326P.

XX PR

PR 04-DEC-2002; 2002US-0429651P.

XX PR

PR 04-DEC-2002; 2002US-0430645P.

XX PR

PR 04-DEC-2002; 2002US-0430657P.

XX PR

PR 04-DEC-2002; 2002US-0430663P.

PR 03-JAN-2003; 2003US-0437914P.  
 PR 17-JAN-2003; 2003US-0440820P.  
 PR 17-JAN-2003; 2003US-0440821P.  
 PR 18-APR-2003; 2003US-0463700P.  
 PR 18-APR-2003; 2003US-0463708P.  
 PR 18-APR-2003; 2003US-0463716P.  
 PR 18-APR-2003; 2003US-0463732P.  
 PR 02-MAY-2003; 2003US-0467199P.  
 PR 02-MAY-2003; 2003US-0467201P.  
 PR 02-MAY-2003; 2003US-0467203P.  
 PR 02-MAY-2003; 2003US-0467230P.  
 PR 19-MAY-2003; 2003US-0471306P.  
 PR 19-MAY-2003; 2003US-0471336P.  
 PR 22-MAY-2003; 2003US-0472420P.  
 PR 22-MAY-2003; 2003US-0472430P.  
 PR 09-JUN-2003; 2003US-0476609P.  
 PR 09-JUN-2003; 2003US-0476621P.  
 PR 09-JUN-2003; 2003US-0476632P.  
 PR 09-JUN-2003; 2003US-0476641P.  
 PR 08-JUL-2003; 2003US-0485217P.  
 PR 08-JUL-2003; 2003US-0485218P.  
 PR 08-JUL-2003; 2003US-0485223P.  
 PR 08-JUL-2003; 2003US-0485224P.  
 PR 08-JUL-2003; 2003US-0485252P.  
 PR 08-JUL-2003; 2003US-0485359P.  
 PR 14-JUL-2003; 2003US-0488448P.  
 PR 14-JUL-2003; 2003US-0488480P.  
 PR 15-JUL-2003; 2003US-0486891P.  
 PR 15-JUL-2003; 2003US-0486960P.  
 PR 08-AUG-2003; 2003US-0493341P.  
 PR 08-AUG-2003; 2003US-0493370P.  
 PR 08-AUG-2003; 2003US-0493573P.  
 PR 08-AUG-2003; 2003US-0493577P.  
 XX XX

(FIVE-) FIVE PRIME THERAPEUTICS INC.

Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;

Halenbeck RF, Kothakota S, Lin H, Linnemann T, Pierce K, Wang Y;

Wong JGP, Wu G, Zhang H, Zeng C;

WPI; 2004-365511/34.

P-PSDB; ADN99956.

New nucleic acid molecules, useful in preparing a composition for treating or preventing e.g. inflammatory, CNS, bacterial or viral disorders, cancer, psoriasis, diabetes, ischemic heart disease or ulcerative colitis.

Claim 1; SEQ ID NO 2340; 532pp; English.

The invention relates to a nucleic acid molecule comprising a polynucleotide sequence or its complement that encodes a polypeptide. The nucleic acid is useful in preparing a composition for treating or preventing inflammatory, CNS, immune, bacterial or viral disorder, cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic heart disease or ulcerative colitis. This sequence corresponds to a nucleic acid of the invention.

Sequence 1287 BP; 451 A; 235 C; 275 G; 326 T; 0 U; 0 Other;

Query Match 78.3%; Score 18; DB 12; Length 1287;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TCTTAAGGAGACACCACA 21

|||||

Db 59-TCTTAAGGAGACACCACA 76

#### RESULT 5

ADN99172

ID ADN99172 standard; cDNA; 1287 BP.

XX XX

```
AC ADN99172;
XX
XX 29-JUL-2004 (first entry)
XX
XX Novel human cDNA sequence #772.
XX
XX ds; gene; anti-inflammatory; dermatological; neuroprotective;
XX immunomodulator; antibacterial; virucide; antiparasitic; cytostatic;
XX gene therapy; vaccine; inflammatory; CNS; immune disorder; cancer;
XX psoriasis; diabetes; early aging; hormonal imbalance;
XX ischemic heart disease; ulcerative colitis.
XX
XX Homo sapiens.
XX
XX WO2004038003-A2.
XX
XX 06-MAY-2004.
XX
XX 24-OCT-2003; 2003WO-US033947.
XX
XX 25-OCT-2002; 2002US-0421061P.
XX 25-OCT-2002; 2002US-0421080P.
XX 25-OCT-2002; 2002US-0421552P.
XX 25-OCT-2002; 2002US-0421614P.
XX 30-OCT-2002; 2002US-0422177P.
XX 30-OCT-2002; 2002US-0422178P.
XX 30-OCT-2002; 2002US-0422179P.
XX 15-NOV-2002; 2002US-0426355P.
XX 15-NOV-2002; 2002US-0426384P.
XX 15-NOV-2002; 2002US-0426394P.
XX 15-NOV-2002; 2002US-0426430P.
XX 15-NOV-2002; 2002US-0426916P.
XX 27-NOV-2002; 2002US-0429224P.
XX 27-NOV-2002; 2002US-0429275P.
XX 27-NOV-2002; 2002US-0429302P.
XX 27-NOV-2002; 2002US-0429326P.
XX 27-NOV-2002; 2002US-0429651P.
XX 04-DEC-2002; 2002US-0430645P.
XX 04-DEC-2002; 2002US-0430651P.
XX 04-DEC-2002; 2002US-0430657P.
XX 04-DEC-2002; 2002US-0430663P.
XX 04-DEC-2002; 2002US-0430668P.
XX 04-DEC-2002; 2002US-0430684P.
XX 05-DEC-2002; 2002US-0430937P.
XX 05-DEC-2002; 2002US-0430965P.
XX 05-DEC-2002; 2002US-0431458P.
XX 12-DEC-2002; 2002US-0433251P.
XX 12-DEC-2002; 2002US-0433500P.
XX 13-DEC-2002; 2002US-0433316P.
XX 13-DEC-2002; 2002US-0433318P.
XX 23-DEC-2002; 2002US-0436236P.
XX 03-JAN-2003; 2003US-0437914P.
XX 17-JAN-2003; 2003US-0440820P.
XX 17-JAN-2003; 2003US-0440821P.
XX 18-APR-2003; 2003US-0463700P.
XX 18-APR-2003; 2003US-0463708P.
XX 18-APR-2003; 2003US-0463716P.
XX 18-APR-2003; 2003US-0463732P.
XX 02-MAY-2003; 2003US-0467199P.
XX 02-MAY-2003; 2003US-0467201P.
XX 02-MAY-2003; 2003US-0467203P.
XX 02-MAY-2003; 2003US-0467230P.
XX 19-MAY-2003; 2003US-0471306P.
XX 22-MAY-2003; 2003US-0472420P.
XX 22-MAY-2003; 2003US-0472430P.
XX 09-JUN-2003; 2003US-0476609P.
XX 09-JUN-2003; 2003US-0476621P.
XX 09-JUN-2003; 2003US-0476632P.
XX 09-JUN-2003; 2003US-0476641P.
XX 08-JUL-2003; 2003US-0485217P.
XX 08-JUL-2003; 2003US-0485218P.
XX 08-JUL-2003; 2003US-0485223P.
XX 08-JUL-2003; 2003US-0485224P.
XX
XX PR 08-JUL-2003; 2003US-0485325P.
XX PR 08-JUL-2003; 2003US-0485359P.
XX PR 14-JUL-2003; 2003US-0486466P.
XX PR 14-JUL-2003; 2003US-0486480P.
XX PR 15-JUL-2003; 2003US-0486891P.
XX PR 15-JUL-2003; 2003US-0486960P.
XX PR 08-AUG-2003; 2003US-0493341P.
XX PR 08-AUG-2003; 2003US-0493370P.
XX PR 08-AUG-2003; 2003US-0493573P.
XX PR 08-AUG-2003; 2003US-0493577P.
XX
XX (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
XX Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;
XX Halenbeck RP, Kothakota S, Lin H, Linnemann T, Pierce K, Wang Y;
XX Wong JGP, Wu G, Zhang H, Zeng C;
XX
XX WPI; 2004-365511/34.
XX PR P-PSDB; ADN99956.
XX
XX New nucleic acid molecules, useful in preparing a composition for
XX treating or preventing e.g. inflammatory, CNS, bacterial or viral
XX disorders, cancer, psoriasis, diabetes, ischemic heart disease or
XX ulcerative colitis.
XX
XX Claim 1; SEQ ID NO 772; 532pp; English.
XX
XX The invention relates to a nucleic acid molecule comprising a
XX polynucleotide sequence or its complement that encodes a polypeptide. The
XX nucleic acid is useful in preparing a composition for treating or
XX preventing inflammatory, CNS, immune, bacterial or viral disorder,
XX cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic
XX heart disease or ulcerative colitis. This sequence corresponds to a
XX nucleic acid of the invention.
XX
XX SQ Sequence 1287 BP; 451 A; 235 C; 275 G; 326 T; 0 U; 0 Other;
XX
XX Query Match 78.3%; Score 18; DB 12; Length 1287;
XX Best Local Similarity 100.0%; Pred. No. 1.5e+02;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 4 TCTTAAGGAGACACCACA 21
XX |||||
XX Db 59 TCTTAAGGAGACACCACA 76
XX
XX RESULT 6
XX AAS72172
XX ID AAS72172 standard; cDNA; 1726 BP.
XX
XX AC AAS72172;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE DNA encoding novel human diagnostic protein #7976.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
```

XX WPI; 2001-639362/73.  
 DR P-PSDB; ABG07985.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 1; SEQ ID NO 7976; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have application in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
 CC coding sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 1726 BP; 602 A; 326 C; 349 G; 449 T; 0 U; 0 Other;  
 Query Match 78.3%; Score 18; DB 5; Length 1726;  
 Best Local Similarity 100.0%; Pred. NO. 1.5e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 TCTTAAGGAGACACCACA 21  
 DB 59 TCTTAAGGAGACACCACA 76  
 RESULT 7  
 ACH87437  
 ID ACH87437 standard; DNA; 1766 BP.  
 AC ACH87437;  
 XX  
 DT 29-JUL-2004 (first entry)  
 DE Human genome derived single exon probe #20632.  
 XX  
 KW Human; probe; ss; gene expression; single exon probe; microarray;  
 KW alternative splicing event; genomic alteration.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2003194704-A1.  
 XX  
 PD 16-OCT-2003.  
 XX  
 PF 03-APR-2002; 2002US-00029386.  
 XX  
 PR 03-APR-2002; 2002US-00029386.  
 XX  
 PA (PENN/) PENN S G.  
 PA (RANK/) RANK D R.  
 PA (HANZ/) HANZEL D K.  
 XX  
 PI Penn SG, Rank DR, Hanzel DK;  
 XX

DR WPI; 2004-119264/12.  
 XX  
 PT New human genome-derived single exon nucleic acid probes useful for human  
 PT gene expression analysis, for identifying or characterizing alternative  
 PT splicing events, for assessing genomic alterations or as tools for  
 PT surveying tissues.  
 XX  
 PS Claim 1; SEQ ID NO 20632; 80pp; English.  
 XX  
 CC The invention relates to a nucleic acid probe for measuring human gene  
 CC expression, comprising any of the 27,400 fully defined nucleotide  
 CC sequences in the specification, or their complements or fragments, and  
 CC encoding at least 8 amino acids of any of the 688 amino acid sequences  
 CC fully defined in the specification. The probe is a single exon probe that  
 CC hybridises under high stringency conditions to a nucleic acid molecule  
 CC expressed in human cells or tissues. Also included are a spatially-  
 CC addressable set of single exon nucleic acid probes for measuring human  
 CC gene expression (comprising a plurality of single exon nucleic acid  
 CC probes cited above, where each of the plurality of probes is separately  
 CC and addressably isolatable or amplifiable from the plurality), a single  
 CC exon microarray for measuring human gene expression, a method of  
 CC measuring human gene expression, a vector comprising the single exon  
 CC probe cited above, an ORF-encoded peptide comprising at least 8  
 CC contiguous amino acids of any of the above-mentioned amino acid  
 CC sequences (optionally with conservative amino acid substitutions), an  
 CC isolated antibody that binds specifically to a peptide cited above,  
 CC methods of selling and/or licensing single exon probes or microarrays to  
 CC a customer desiring to measure gene expression, a method of providing  
 CC human gene expression data by subcription, and a computer-readable  
 CC storage medium which contains a database having a plurality of records  
 CC (each record including data on the expression of a single exon probe  
 CC cited above. The probe, methods and apparatus are useful in gene  
 CC expression analysis. The probes may be used as tools for surveying  
 CC tissues to detect the presence of expressed messages that contain their  
 CC specific exon, or in constructing genome-derived single exon microarrays.  
 CC In addition, the probes are used in identifying and characterising  
 CC alternative splicing events, in detecting and characterising gross  
 CC alterations in the genomic locus that includes their exon, in assessing  
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
 CC or in expressing the ORF-encoded peptide. The present sequence is a human  
 CC single exon probe of the invention. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?DocID=20030194704  
 XX  
 SQ Sequence 1766 BP; 619 A; 331 C; 366 G; 450 T; 0 U; 0 Other;  
 Query Match 78.3%; Score 18; DB 12; Length 1766;  
 Best Local Similarity 100.0%; Pred. NO. 1.5e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 TCTTAAGGAGACACCACA 21  
 DB 538 TCTTAAGGAGACACCACA 555  
 RESULT 8  
 AAZ52430  
 ID AAZ52430 standard; DNA; 2077 BP.  
 XX  
 AC AAZ52430;  
 XX  
 DT 24-FEB-2000 (first entry)  
 XX  
 DE HTRM clone 438283 DNA sequence.  
 XX  
 KW HTRM; human transcriptional regulatory molecule; arteriosclerosis; AIDS;  
 KW arteriosclerosis; cirrhosis; cancer; leukaemia; diabetes mellitus; ss;  
 KW Addison's disease; multiple sclerosis; rheumatoid arthritis; infection;  
 KW trauma; myasthenia gravis; adenocarcinoma; immune disorder; treatment.  
 XX  
 OS Homo sapiens.  
 XX

PN WO9957144-A2.  
 XX 11-NOV-1999.  
 XX 04-MAY-1999; 99WO-US009935.  
 XX 05-MAY-1998; 98US-0084254P.  
 PR 07-AUG-1998; 98US-0095827P.  
 PR 02-OCT-1998; 98US-0102745P.  
 XX  
 PA (INCY-) INCYTE PHARM INC.  
 XX Hillman JL, Bandman O, Lal P, Yue H, Reddy R, Tang YT;  
 PI Gerstin EH, Patterson C, Baughn MR, Azimzai Y, Lu DAM;  
 XX WPI; 2000-052941/04.  
 DR P-PSDB; AAY73345.  
 XX  
 PT New peptides useful for diagnosis, prevention and treatment of cancer and  
 PT immune disorders.  
 XX  
 PS Claim 9; Page 160; 193pp; English.  
 XX  
 CC AAZ52410-Z52474 are human transcriptional regulator molecule (HTRM)  
 CC nucleotide sequences. The HTRM protein and nucleotide sequences are  
 CC useful for preventing or treating disorders associated with decreased  
 CC expression or activity of HTRM which include cell proliferative disorders  
 CC such as arteriosclerosis and cirrhosis; cancers including adenocarcinoma  
 CC and leukaemia; immune disorders such as AIDS, Addison's disease, diabetes  
 CC mellitus, rheumatoid arthritis, multiple sclerosis, systemic lupus  
 CC erythematosus, and myasthenia gravis; infections and trauma. Antagonists  
 CC of the HTRM polypeptides are useful for treating or preventing disorders  
 CC associated with increased expression or activity of HTRMs. HTRM  
 CC polypeptides, their immunogenic fragments or oligopeptides are useful for  
 CC screening libraries of compounds in drug screening techniques.  
 CC Polynucleotides encoding HTRM are useful for blocking the transcription  
 CC of mRNA and regulating gene function by modulating the activity of HTRM.  
 CC Vectors expressing HTRM or agonists can also be used to prevent or treat  
 CC disorder associated with decreased HTRM expression. Antibodies which  
 CC specifically bind HTRM and polynucleotides encoding HTRM are useful for  
 CC diagnosing disorders associated with the expression of HTRM, particularly  
 CC in assays that detect the expression of HTRM. Nucleotide sequences  
 CC encoding HTRM may be useful to generate hybridization probes useful in  
 CC mapping the naturally occurring genomic sequence and to detect  
 CC differences in gene sequences among normal, carrier and affected  
 CC individuals. Using diagnostic assays, cancer can be detected prior to the  
 CC appearance of clinical symptoms and thereby progression of cancer can be  
 CC prevented by aggressive treatment or preventive measures  
 XX  
 SQ Sequence 2077 BP; 686 A; 405 C; 456 G; 530 T; 0 U; 0 Other;  
 Query Match 78.3%; Score 18; DB 3; Length 2077;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 TCTTAGGAGACACACCA 21  
 Db 885 TCTTAGGAGACACACCA 902  
 RESULT 9  
 ID ABX15830/c  
 XX ABX15830 standard; cDNA; 1646 BP.  
 AC ABX15830;  
 XX 31-MAR-2003 (first entry)  
 XX cDNA encoding soybean sterol delta-7 reductase protein #2.  
 DE Corn; wheat; rice; soybean; plant; gene; ss; sterol metabolism;  
 XX sterol delta 7; reductase; sterol C5 desaturase.  
 KW  
 XX

OS Glycine max.  
 XX US6465717-B1.  
 XX 15-OCT-2002.  
 XX 19-NOV-1999; 99US-00443041.  
 XX 20-NOV-1998; 98US-0109283P.  
 XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX Famodu OO, Rafalski JA;  
 XX WPI; 2003-147075/14.  
 DR P-PSDB; ABG73986.  
 XX  
 PT Novel polynucleotide encoding sterol metabolism enzyme useful for  
 PT producing transgenic plant, has nucleotide sequence encoding the  
 PT polypeptide having sterol delta-7 reductase activity.  
 XX  
 PS Claim 5; Col 43-46; 39pp; English.  
 XX  
 CC This invention relates to novel isolated polynucleotide encoding sterol  
 CC metabolism enzyme. The nucleotide sequence of the invention is useful for  
 CC transforming a cell by introducing the sequence into a cell. The  
 CC sequences may also be useful for producing a transgenic plant, by  
 CC transforming a plant cell with the cDNA sequence and regenerating a  
 CC transgenic plant from the transformed plant cell. The nucleotide sequence  
 CC is useful for obtaining a nucleic acid fragment encoding all or a  
 CC substantial portion of the amino acid sequence encoding a sterol delta-7  
 CC reductase or a sterol-C5 desaturase polypeptide. Probes derived from the  
 CC nucleotide sequence of the invention may be useful for physical mapping.  
 CC The present sequence represents a cDNA sequence encoding a sterol  
 CC metabolism enzyme of the invention  
 XX  
 SQ Sequence 1646 BP; 444 A; 343 C; 350 G; 509 T; 0 U; 0 Other;  
 Query Match 77.4%; Score 17.8; DB 8; Length 1646;  
 Best Local Similarity 90.5%; Pred. No. 1.9e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GTGTCTTAAGGAGACACCA 21  
 Db 1460 GTGTCTTAAGGAGACCAAA 1440  
 RESULT 10  
 ID ADI18998/c  
 XX ADI18998 standard; cDNA; 1646 BP.  
 AC ADI18998;  
 XX 15-APR-2004 (first entry)  
 XX Soybean sterol delta-7 reductase cDNA #2.  
 DE Transcription factor; genetic mapping; plant breeding;  
 XX sterol delta-7 reductase; soybean; gene; ss.  
 KW Glycine max.  
 XX Key Location/Qualifiers  
 FT CDS 59..1357  
 FT /\*tag= a  
 FT /product= "sterol delta-7 reductase"  
 XX US2003229914-A1.  
 XX 11-DEC-2003.  
 XX 21-AUG-2002; 2002US-00224880.  
 XX

PR 20-NOV-1998; 98US-0109283P.  
PR 19-NOV-1999; 99US-00443041.  
XX (PAMO/) FAMODU O O.  
PA (OROZ/) OROZCO E M.  
PA (RAFA/) RAFALSKI J A.  
PA (SHEN/) SHEN J B.  
XX  
PI Famodu OO, Orozco EM, Rafalski JA, Shen JB;  
XX WPI; 2004-052051/05.  
DR P-PSDB; AD118999.  
XX  
PT New polynucleotide encoding a transcription factor polypeptide, useful in  
genetic mapping for plant breeding.  
XX Claim 2; SEQ ID NO 19; 43pp; English.  
XX  
CC The invention relates to a polynucleotide which encodes a transcription  
factor. The nucleic acid fragment is useful in genetic mapping for plant  
breeding. The present sequence is sterol delta-7 reductase cDNA.  
XX  
SQ Sequence 1646 BP; 444 A; 343 C; 350 G; 509 T; 0 U; 0 Other;  
Query Match 77.4%; Score 17.8; DB 12; Length 1646;  
Best Local Similarity 90.5%; Pred. No. 1.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GTGCTTAAAGGACACCA 21  
DB 1460 GTGCTTAAAGGACCA 1440  
RESULT 11  
ADA02738/c  
ID ADA02738 standard; DNA; 52302 BP.  
XX  
AC ADA02738;  
XX  
DT 06-NOV-2003 (first entry)  
XX  
DE Human CCND2 carcinoma associated gene, SEQ ID NO:1256.  
XX  
KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;  
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
KW gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO2003057146-A2.  
XX  
PD 17-JUL-2003.  
XX  
PF 26-DEC-2002; 2002WO-US041414.  
XX  
PR 26-DEC-2001; 2001US-00035832.  
XX  
PA (SAGR-) SAGRES DISCOVERY.  
XX  
PI Morris DW;  
XX  
DR WPI; 2003-587068/55.  
XX  
PT New recombinant nucleic acid encoding carcinoma associated protein,  
PT useful for preparing compositions for treating carcinomas.  
XX  
PS Claim 1; SEQ ID NO 1256; 245pp; English.  
XX  
CC The invention relates to recombinant carcinoma associated (CA) nucleic  
acid sequences from mouse and human (ADA01482-ADA03094), and to  
CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
CC invention also encompasses expression vectors and host cells comprising a  
CC nucleic acid, a polypeptide (especially an antibody) that specifically

CC binds to the protein, and a biochip comprising CA nucleic acid or  
CC fragments thereof. The sequences of the invention were identified using  
CC oncogenic retroviruses, which insert into the genome of the host organism  
CC at random. Many of these do not carry transduced host oncogenes or  
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
CC direct consequence of the effects of proviral integration into host  
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
CC leukaemia) or a propensity to carcinoma by determination of the sequence  
CC of a CA gene, or by determination of CA gene expression in particular  
CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
CC therapeutic agents and in screening and evaluating drug candidates. The  
CC present sequence represents a specifically claimed human CA nucleic acid  
CC sequence of the invention. Note: The complete sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;  
Query Match 77.4%; Score 17.8; DB 9; Length 52302;  
Best Local Similarity 90.5%; Pred. No. 2.8e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 3 GTCTTAAAGGACACCA 23  
DB 13323 GTTTTAGGAGACCA 13303  
RESULT 12  
ADB72476/c  
ID ADB72476 standard; DNA; 52302 BP.  
XX  
AC ADB72476;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Human CCND2 gene.  
XX  
KW human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;  
KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.  
XX  
OS Homo sapiens.  
XX  
PN WO2003008583-A2.  
XX  
PD 30-JAN-2003.  
XX  
PF 26-DEC-2001; 2001WO-US051291.  
XX  
PR 02-MAR-2001; 2001US-00798586.  
PR 23-OCT-2001; 2001US-00004113.  
PR 08-NOV-2001; 2001US-00052482.  
PR 30-NOV-2001; 2001US-00997722.  
PR 20-DEC-2001; 2001US-00034650.  
XX  
PA (SAGR-) SAGRES DISCOVERY.  
XX  
PI Morris DW, Engelhard EK;  
XX  
DR WPI; 2003-239337/23.  
XX  
PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,  
PT cancers, neoplasm, adenocarcinoma, or sarcomas.  
XX  
PS Claim 1; SEQ ID NO 304; 2304pp; English.  
XX  
CC The invention relates to a novel recombinant nucleic acid comprising a  
CC nucleotide sequence selected from any of the 660 sequences fully defined  
CC in the specification. A polynucleotide of the invention has cytostatic  
CC activity, and may have a use in gene therapy, or in a vaccine. The  
CC recombinant nucleic acids and polypeptides are useful for treating  
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and

```
CC sarcomas. The present sequence represents a human gene of the invention.
XX
SQ Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

Query Match          77.4%; Score 17.8; DB 10; Length 52302;
Best Local Similarity 90.5%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 GTCTTAAGGAGACACACAGC 23
Db      13323 GTTTTAGGAGACACACAGC 13303

RESULT 13
ID ADC85218/c
XX ADC85218 standard; DNA; 52302 BP.
AC ADC85218;
XX
DT 01-JAN-2004 (first entry)
XX Human Ccnd2 genomic sequence.
DE
XX Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
KW secreted; transmembrane; intracellular; ds.
XX Homo sapiens.
OS
XX WO2003045230-A2.
PN
XX 05-JUN-2003.
XX
XX 02-DEC-2002; 2002WO-US038582.
XX
XX 30-NOV-2001; 2001US-00997722.
XX
XX (SAGR-) SAGRES DISCOVERY.
XX
XX Morris DW, Engelhard EK;
XX WPI; 2003-513603/48.
XX
XX New recombinant nucleic acid comprising a nucleotide sequence of any of
PT the carcinoma-associated (CA) genes, useful for screening for drug
PT candidates for diagnosing or treating carcinomas.
XX
XX Claim 1; SEQ ID NO 4; 983pp; English.
XX
XX The invention relates to a recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the fully defined carcinoma-
CC associated (CA) genes from the 50 tables given in the specification. The
CC CA proteins are secreted, transmembrane or intracellular proteins. The
CC recombinant nucleic acids are useful for screening for drug candidates
CC for diagnosing or treating carcinomas. Sequences given in ADC85215-
XX ADC85514 represent CA genes of the invention.
XX
SQ Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

Query Match          77.4%; Score 17.8; DB 10; Length 52302;
Best Local Similarity 90.5%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 GTCTTAAGGAGACACACAGC 23
Db      13323 GTTTTAGGAGACACACAGC 13303

RESULT 14
ID ADM74333/c
XX ADM74333 standard; DNA; 52302 BP.
XX
XX ADM74333;
XX
```

```
DT 01-JUL-2004 (first entry)
XX
XX Human carcinoma associated (CA) nucleic acid #1.
XX
XX Human; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;
KW carcinoma associated protein; CAP; carcinoma; leukaemia; lymphoma;
KW cytostatic.
XX
XX Homo sapiens.
XX
XX US2004072154-A1.
PN
XX 15-APR-2004.
XX
XX 30-NOV-2001; 2001US-00997722.
XX
XX 22-DEC-2000; 2000US-00747377.
PR
XX 02-MAR-2001; 2001US-00798586.
XX
XX (MORR/) MORRIS D W.
XX (ENG/) ENGELHARD E K.
XX
XX Morris DW, Engelhard EK;
PI
XX WPI; 2004-328562/30.
XX
XX New carcinoma associated gene or protein, useful for preparing a
PT composition for diagnosing or treating carcinoma e.g., leukemia or
PT lymphoma.
XX
XX Claim 1; SEQ ID NO 4; 29pp; English.
XX
XX The invention relates to new recombinant nucleic acids. The invention
CC also relates to a host cell comprising a recombinant nucleic acid or
CC expression vector, an expression vector comprising a recombinant nucleic
CC acid, a recombinant protein, a method of screening for drug candidates, a
CC method of screening for a bioactive agent capable of binding to a
CC carcinoma associated protein (CAP) encoded by a nucleotide sequence, a
CC method of screening for a bioactive agent capable of modulating the
CC activity of a CAP, a method of evaluating the effect of a candidate
CC carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting
CC the activity of a CAP, a method of treating carcinomas, a method of
CC neutralising the effect of a CAP and a method of diagnosing carcinoma or
CC propensity to carcinoma. A method of evaluating the effect of a candidate
CC carcinoma drug comprises administering the drug to a patient, removing a
CC cell sample from the patient and determining alterations in the
CC expression or activation of a gene comprising the nucleotide sequence. A
CC method of diagnosing carcinoma comprises determining the expression of
CC one or more genes comprising the nucleic acid sequence in a first tissue
CC type of a first individual and comparing the expression of the gene from
CC a second normal tissue type from the first individual or a second
CC unaffected individual, where a difference in the expression indicates
CC that the first individual has carcinoma. A method of inhibiting the
CC activity of a CAP comprises binding an inhibitor to the CAP. Treating
CC carcinomas comprises administering to a patient an inhibitor of CAP.
CC Neutralising the effect of a CAP comprises contacting an agent specific
CC for the CAP. The polypeptide specifically binds to the protein encoded by
CC the nucleic acid. It comprises an antibody that specifically binds to the
CC protein encoded by the nucleic acid. The nucleic acids are useful for
CC preparing a composition for diagnosing or treating carcinoma e.g.,
CC leukaemia or lymphoma. This sequence represents a human carcinoma
CC associated (CA) nucleic acid of the invention. Note: The sequence data
CC for this patent did not form part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

Query Match          77.4%; Score 17.8; DB 12; Length 52302;
Best Local Similarity 90.5%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 GTCTTAAGGAGACACACAGC 23
```



Db 13323 GTTTTAGGAGACACACAGC 13303

RESULT 15

ACN44062

ID ACN44062 standard; DNA; 144035 BP.

XX

XX ACN44062;

XX

DT 18-NOV-2004 (first entry)

XX

DE Human genomic sequence hCG38705.

XX

KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.

XX

OS Homo sapiens.

XX

PN WO2003073826-A2.

XX

PD 12-SEP-2003.

XX

PF 28-FEB-2003; 2003WO-US006235.

XX

PR 01-MAR-2002; 2002US-00087192.

XX

PA (SAGR-) SAGRES DISCOVERY.

XX

PI Morris DW;

XX

DR WPI; 2003-328604/31.

XX

PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma comprises a nucleotide sequence.

PT

PS Claim 1; SEQ ID NO 322; Opp; English.

XX

CC The present invention relates to novel DNA and protein sequences which are associated with carcinomas. The sequences are useful for: (i) for screening drug candidates; (ii) for screening of bioactive agent capable of binding to Carcinoma Associated Protein (CAP); (iii) for screening of a bioactive agent capable of modulating the activity of CAP; (iv) for evaluating the effect of a candidate carcinoma drug; (v) for diagnosing carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip; (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for determining Carcinoma Associated (CA) gene copy number. In addition, the CA genes are useful as DNA vaccines and the CAP are useful as markers of carcinoma including lymphoma. The present sequence is one such CA coding sequence. Note: This patent is an equivalent to basic patent US2002182586A1, for which no sequence data was published

XX SQ Sequence 144035 BP; 43371 A; 28888 C; 27578 G; 44198 T; 0 U; 0 Other;

Query Match

Best Local Similarity 77.4%; Score 17.8; DB 11; Length 144035;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 TGTCTTAAGAGACACACAG 22

Db 137647 TGTCTTAAGAGACACACAG 137667

Search completed: March 9, 2005, 07:25:24  
Job time : 430 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2005, 05:50:50 ; Search time 127 Seconds  
(without alignments)  
296.334 Million cell updates/sec

Title: US-09-835-976b-115

Perfect score: 23

Sequence: 1 ggtcttaaggagaccacacagc 23

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

- 1: /cgn2.6/prodata/1/ina/5A COMB.seq.\*
- 2: /cgn2.6/prodata/1/ina/5B COMB.seq.\*
- 3: /cgn2.6/prodata/1/ina/6A COMB.seq.\*
- 4: /cgn2.6/prodata/1/ina/6B COMB.seq.\*
- 5: /cgn2.6/prodata/1/ina/PCUS COMB.seq.\*
- 6: /cgn2.6/prodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18.8	81.7	102053	4	US-09-949-016-13025
2	17.8	77.4	601	4	US-09-949-016-23241
3	17.8	77.4	601	4	US-09-949-016-73745
4	17.8	77.4	1646	3	US-09-443-041A-19
5	17.8	77.4	36302	4	US-09-949-016-11998
6	17.8	77.4	36302	4	US-09-949-016-13891
7	17.2	74.8	1603	3	US-09-103-664A-3
8	17.2	74.8	1889	3	US-09-103-664A-10
9	17.2	74.8	2091	3	US-09-103-664A-1
10	17.2	74.8	2872	1	US-08-015-985-4
11	17.2	74.8	2872	4	US-09-280-597-4
12	16.8	73.0	25377	4	US-09-949-016-14260
13	16.8	73.0	25377	4	US-09-949-016-14261
14	16.8	73.0	28791	4	US-09-949-016-15396
15	16.6	72.2	1605	4	US-09-023-655-194
16	16.6	72.2	49407	4	US-09-949-016-12532
17	16.6	72.2	49408	4	US-09-949-016-17045
18	16.6	72.2	301828	4	US-09-949-016-13869
19	16.4	71.3	601	4	US-09-949-016-29750
20	16.4	71.3	601	4	US-09-949-016-63179
21	16.4	71.3	38239	4	US-09-949-016-12348
22	16.4	71.3	38252	4	US-09-949-016-13570
23	16.4	71.3	670889	4	US-09-949-016-12505
24	16.4	71.3	670690	4	US-09-949-016-14207
25	16.2	70.4	25	4	US-09-396-196G-42041
26	16.2	70.4	25	4	US-09-396-196G-42042
27	16.2	70.4	25	4	US-09-396-196G-42044

```

28 16.2 70.4 564 4 US-09-248-796A-5566 Sequence 5566, Ap
29 16.2 70.4 601 4 US-09-949-016-203490 Sequence 203490,
30 16.2 70.4 601 4 US-09-949-016-203491 Sequence 203491,
31 16.2 70.4 601 4 US-09-949-016-203492 Sequence 203492,
32 16.2 70.4 56976 4 US-09-949-016-17486 Sequence 17486, A
33 16.2 70.4 75176 4 US-09-949-016-13300 Sequence 13300, A
34 16.2 70.4 101574 4 US-09-949-016-12033 Sequence 12033, A
35 16.2 70.4 169998 3 US-09-676-610B-24 Sequence 24, Appl
36 16.2 70.4 197496 4 US-09-877-177A-10 Sequence 10, Appl
37 16.2 69.6 2107 4 US-09-774-528-332 Sequence 332, Appl
38 15.8 68.7 601 4 US-09-949-016-53600 Sequence 53600, A
39 15.8 68.7 601 4 US-09-949-016-77637 Sequence 77637, A
40 15.8 68.7 601 4 US-09-949-016-77638 Sequence 77638, A
41 15.8 68.7 601 4 US-09-949-016-132754 Sequence 132754,
42 15.8 68.7 601 4 US-09-949-016-132755 Sequence 132755,
43 15.8 68.7 601 4 US-09-949-016-138467 Sequence 138467,
44 15.8 68.7 601 4 US-09-949-016-175717 Sequence 175717,
45 15.8 68.7 601 4 US-09-949-016-194492 Sequence 194492,

```

#### ALIGNMENTS

##### RESULT 1

```

US-09-949-016-13025
; Sequence 13025, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13025
; LENGTH: 102053
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(102053)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13025

```

```

Query Match 81.7%; Score 18.8; DB 4; Length 102053;
Best Local Similarity 90.9%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 2 TGCTTTAAGGAGACACACAGC 23

Db 87025 TGCTTTAAGGAGACACATCAGC 87046

##### RESULT 2

```

US-09-949-016-23241/C
; Sequence 23241, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755

```



; LENGTH: 36302  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(36302)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-13891

Query Match 77.4%; Score 17.8; DB 4; Length 36302;  
Best Local Similarity 90.5%; Pred. No. 64;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GTCTTAAGGAGACACACAGC 23  
DB 5323 GTTTTAGGAGACACACAGC 5303

## RESULT 7

US-09-103-664A-3/c  
; Sequence 3, Application US/09103664A  
; Patent No. 6458358  
; GENERAL INFORMATION:  
; APPLICANT: University of Kentucky Research Foundation  
; APPLICANT: Timoney, John  
; APPLICANT: Artushin, Sergey  
; TITLE OF INVENTION: Compounds Encoding the Protective M-Like Protein of Streptococcus  
; FILE OF INVENTION: and Assays Therefor  
; FILE REFERENCE: 50229-212  
; CURRENT APPLICATION NUMBER: US/09/103,664A  
; CURRENT FILING DATE: 1998-06-23  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 3  
; LENGTH: 1603  
; TYPE: DNA  
; ORGANISM: Streptococcus equi  
US-09-103-664A-3

Query Match 74.8%; Score 17.2; DB 3; Length 1603;  
Best Local Similarity 86.4%; Pred. No. 85;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TGCTTAAGGAGACACACAGC 23  
DB 1594 TCGTTTAGGAGACACACAGC 1573

## RESULT 8

US-09-103-664A-10/c  
; Sequence 10, Application US/09103664A  
; Patent No. 6458358  
; GENERAL INFORMATION:  
; APPLICANT: University of Kentucky Research Foundation  
; APPLICANT: Timoney, John  
; APPLICANT: Artushin, Sergey  
; TITLE OF INVENTION: Compounds Encoding the Protective M-Like Protein of Streptococcus  
; FILE OF INVENTION: and Assays Therefor  
; FILE REFERENCE: 50229-212  
; CURRENT APPLICATION NUMBER: US/09/103,664A  
; CURRENT FILING DATE: 1998-06-23  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 10  
; LENGTH: 1889  
; TYPE: DNA  
; ORGANISM: Streptococcus equi  
US-09-103-664A-10

Query Match 74.8%; Score 17.2; DB 3; Length 1889;  
Best Local Similarity 86.4%; Pred. No. 87;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TGCTTAAGGAGACACACAGC 23  
DB 1732 TCGTTTAGGAGACACACAGC 1711

## RESULT 9

US-09-103-664A-1/c  
; Sequence 1, Application US/09103664A  
; Patent No. 6458358  
; GENERAL INFORMATION:  
; APPLICANT: University of Kentucky Research Foundation  
; APPLICANT: Timoney, John  
; APPLICANT: Artushin, Sergey  
; TITLE OF INVENTION: Compounds Encoding the Protective M-Like Protein of Streptococcus  
; FILE OF INVENTION: and Assays Therefor  
; FILE REFERENCE: 50229-212  
; CURRENT APPLICATION NUMBER: US/09/103,664A  
; CURRENT FILING DATE: 1998-06-23  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 2091  
; TYPE: DNA  
; ORGANISM: Streptococcus equi  
US-09-103-664A-1

Query Match 74.8%; Score 17.2; DB 3; Length 2091;  
Best Local Similarity 86.4%; Pred. No. 88;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TGCTTAAGGAGACACACAGC 23  
DB 1934 TCGTTTAGGAGACACACAGC 1913

## RESULT 10

US-08-015-985-4  
; Sequence 4, Application US/08015985  
; Patent No. 5538886  
; GENERAL INFORMATION:  
; APPLICANT: Schlesinger, Joseph  
; APPLICANT: Sap, Jan M.  
; TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE  
; FILE OF INVENTION: PHOSPHATASE-ALPHA  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 AVENUE OF THE AMERICAS  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/015,985  
; FILING DATE: 10-FEB-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7683-020  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2872 base pairs  
; TYPE: nucleic acid

```
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-015-985-4

Query Match      74.8%; Score 17.2; DB 1; Length 2872;
Best Local Similarity 86.4%; Pred. No. 92;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GTGCTTTAAGGAGACACACAC 22
Db      2188 GTGCTCTACGGAGACATCAG 2209

RESULT 11
US-09-280-597-4
; Sequence 4, Application US/09280597
; Patent No. 6682905
; GENERAL INFORMATION:
; APPLICANT: Schlensing, Joseph
; APPLICANT: Sap, Jan M.
; TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
; TITLE OF INVENTION: PHOSPHATASE-ALPHA
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/280,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/015,985
; FILING DATE: 10-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7683-020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2872 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-09-280-597-4

Query Match      74.8%; Score 17.2; DB 4; Length 2872;
Best Local Similarity 86.4%; Pred. No. 92;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GTGCTTTAAGGAGACACACAC 22
Db      2188 GTGCTCTACGGAGACATCAG 2209

RESULT 12
US-09-949-016-14260
; Sequence 14260, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14260
; LENGTH: 25377
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14260

Query Match      73.0%; Score 16.8; DB 4; Length 25377;
Best Local Similarity 90.0%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GTGCTTTAAGGAGACACACAC 20
Db      12407 GTGCTTCAGGAGACACTAC 12426

RESULT 13
US-09-949-016-14261
; Sequence 14261, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14261
; LENGTH: 25377
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14261

Query Match      73.0%; Score 16.8; DB 4; Length 25377;
Best Local Similarity 90.0%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GTGCTTTAAGGAGACACACAC 20
Db      12407 GTGCTTCAGGAGACACTAC 12426

RESULT 14
US-09-949-016-15396/c
; Sequence 15396, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
```

CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15396  
LENGTH: 28791  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-15396

Query Match 73.0%; Score 16.8; DB 4; Length 28791;  
Best Local Similarity 90.0%; Pred. No. 1.9e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 TGCTTTAAGGAGACACCA 21  
DB 3097 TGCTTTAAGGAGACACCA 3078

RESULT 15  
US-09-023-655-194/c  
Sequence 194, Application US/09023655  
Patent No. 6607879  
GENERAL INFORMATION:  
APPLICANT: Cocks, Benjamin G.  
APPLICANT: Susan G. Stuart  
APPLICANT: Jeffrey J. Seilhamer  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
TITLE OF INVENTION: EXPRESSION  
NUMBER OF SEQUENCES: 1508  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,655  
FILING DATE: HEREWITH  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Zeller, Karen J.  
REGISTRATION NUMBER: 37,071  
REFERENCE/DOCKET NUMBER: PA-0001 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 194:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1605 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: THFIPER01  
CLONE: 073293  
US-09-023-655-194

Query Match 72.2%; Score 16.6; DB 4; Length 1605;  
Best Local Similarity 82.6%; Pred. No. 1.7e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 GTGTCTTTAAGGAGACACCA 23  
DB 533 GTGTCTTTAAGGAGACCAAGC 511  
Search completed: March 9, 2005, 08:49:55  
Job time: 129 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2005, 07:18:21 ; Search time 509 Seconds  
(without alignments)  
268.131 Million cell updates/sec

Title: US-09-835-976B-115

Perfect score: 23

Sequence: 1 GTGTCTTAAGGAGACACCACG 23

Scoring table: IDENTITY NUC

Gapop 10\_0 , Gapext 1.0

Searched: 5401638 seqs, 2966923429 residues

Total number of hits satisfying chosen parameters: 10803276

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*
- 19: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 20: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 21: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 22: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	100.0	23	10	US-09-835-976B-115
2	18	78.3	1766	16	US-10-029-386-20632
3	17.8	77.4	1646	17	US-10-024-880C-19
4	17.8	77.4	1721	17	US-10-424-599-49392
5	17.8	77.4	52302	11	US-09-997-722-4
6	17.8	77.4	144035	13	US-10-087-192-322
7	17.4	75.7	201	18	US-10-741-601-9071
8	17.4	75.7	201	18	US-10-741-601-9089
9	17.4	75.7	201	18	US-10-741-601-9487
10	17.4	75.7	201	18	US-10-741-601-12984
11	17.4	75.7	201	18	US-10-741-601-12992

12	17.4	75.7	201	18	US-10-741-601-13191	Sequence 13191, A
13	17.4	75.7	201	19	US-10-741-600-23429	Sequence 23429, A
14	17.4	75.7	201	19	US-10-741-600-23447	Sequence 23447, A
15	17.4	75.7	201	19	US-10-741-600-23845	Sequence 23845, A
16	17.4	75.7	201	19	US-10-741-600-31264	Sequence 31264, A
17	17.4	75.7	201	19	US-10-741-600-31272	Sequence 31272, A
18	17.4	75.7	201	19	US-10-741-600-31471	Sequence 31471, A
19	17.4	75.7	1543	17	US-10-424-599-124860	Sequence 124860, A
20	17.4	75.7	112486	18	US-10-741-601-5641	Sequence 5641, Ap
21	17.4	75.7	112486	19	US-10-741-601-17642	Sequence 17642, A
22	17.4	75.7	161700	18	US-10-741-601-5623	Sequence 5623, Ap
23	17.4	75.7	161700	19	US-10-741-600-17590	Sequence 17590, A
24	17.2	74.8	146	17	US-10-210-556-99	Sequence 99, Appl
25	17.2	74.8	400	17	US-10-242-535A-7892	Sequence 7892, Ap
26	17.2	74.8	400	17	US-10-085-783A-7892	Sequence 7892, Ap
27	17.2	74.8	582	13	US-10-027-632-226823	Sequence 226823, A
28	17.2	74.8	582	17	US-10-027-632-226823	Sequence 226823, A
29	17.2	74.8	753	13	US-10-027-632-26433	Sequence 26433, A
30	17.2	74.8	753	17	US-10-027-632-26433	Sequence 26433, A
31	17.2	74.8	2872	19	US-10-777-144-4	Sequence 4, Appl1
32	17.2	74.8	2872	19	US-10-777-145-4	Sequence 4, Appl1
33	17.2	74.8	2872	19	US-10-777-186-4	Sequence 11, Appl
34	17.2	74.8	3070	17	US-10-210-556-11	Sequence 11, Appl
35	17.2	74.8	84073	13	US-10-087-192-712	Sequence 712, App
36	17.2	74.8	202251	13	US-10-087-192-985	Sequence 985, App
37	17.2	74.8	374849	13	US-10-087-192-1627	Sequence 1627, Ap
38	16.8	73.0	417	18	US-10-357-930-17876	Sequence 17876, A
39	16.8	73.0	424	16	US-10-116-712-124	Sequence 124, App
40	16.8	73.0	534	18	US-10-357-930-47694	Sequence 47694, A
41	16.8	73.0	555	13	US-10-027-632-86997	Sequence 86997, A
42	16.8	73.0	555	17	US-10-027-632-86997	Sequence 86997, A
43	16.8	73.0	577	10	US-09-918-995-26549	Sequence 26549, A
44	16.8	73.0	606	13	US-10-027-632-80910	Sequence 80910, A
45	16.8	73.0	606	13	US-10-027-632-80911	Sequence 80911, A

#### ALIGNMENTS

##### RESULT 1

US-09-835-976B-115  
; Sequence 115, Application US/09835976B  
; Publication No. US20030027983A1  
; GENERAL INFORMATION:  
; APPLICANT: Mount, David B.  
; APPLICANT: Delpire, Eric  
; APPLICANT: Gamba, Gerardo  
; APPLICANT: Alfred L. George, Jr.  
; TITLE OF INVENTION: PURIFIED AND ISOLATED POTASSIUM-CHLORIDE COTRANSPORTER NUCLEIC ACJ  
; TITLE OF INVENTION: POLYPEPTIDES AND  
; TITLE OF INVENTION: THERAPEUTIC AND SCREENING METHODS USING SAME  
; FILE REFERENCE: Attorney Docket No. US20030027983A1 1242-26-2  
; CURRENT APPLICATION NUMBER: US/09/835,976B  
; CURRENT FILING DATE: 2001-04-16  
; NUMBER OF SEQ ID NOS: 131  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 115  
; LENGTH: 23  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-835-976B-115

Query Match 100.0%; Score 23; DB 10; Length 23;  
Best Local Similarity 100.0%; Pred No. 0.11; Indels 0; Gaps 0;  
Matches 23; Conservative 0; Mismatches 0;

##### QY 1 GTGTCTTAAGGAGACACCACG 23

|||||  
1 GTGTCTTAAGGAGACACCACG 23

##### RESULT 2

US-10-029-386-20632

; Sequence 20632, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
; FILE REFERENCE: AEOMICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 20632  
; LENGTH: 1766  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC010328.4  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.1  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.9  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.3  
; OTHER INFORMATION: EST HUMAN HIT: BE253605.1, EVALUE 0.00e+00  
; OTHER INFORMATION: SWISSPROT HIT: P51522, EVALUE 0.00e+00  
; OTHER INFORMATION: NT HIT: G18922743, EVALUE 1.00e-128  
US-10-029-386-20632

Query Match 78.3%; Score 18; DB 16; Length 1766;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TCTTAAGGAGACACCA 21  
Db 538 TCTTAAGGAGACACCA 555

## RESULT 3

US-10-224-880C-19/c  
; Sequence 19, Application US/10224880C  
; Publication No. US20030229914A1  
; GENERAL INFORMATION:  
; APPLICANT: Panodu, Omolayo O.  
; APPLICANT: Orozco, Emil M.  
; APPLICANT: Rafalski, J. Antoni  
; APPLICANT: Shen, Jennie B.  
; TITLE OF INVENTION: Sterol Metabolism Enzymes  
; FILE REFERENCE: BB1282 US DIV  
; CURRENT APPLICATION NUMBER: US/10/224,880C  
; CURRENT FILING DATE: 2002-08-21  
; PRIOR APPLICATION NUMBER: 60/109,283  
; PRIOR FILING DATE: 1998-11-20  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 19  
; LENGTH: 1646  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-10-224-880C-19

Query Match 77.4%; Score 17.8; DB 17; Length 1646;  
Best Local Similarity 90.5%; Pred. No. 67;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTGTCTTAAGGAGACACCA 21  
Db 1460 GTGTCTTAAGGAGACCA 1440

## RESULT 4

US-10-424-599-49392/c  
; Sequence 49392, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 49392  
; LENGTH: 1721  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_15608C.1  
US-10-424-599-49392

Query Match 77.4%; Score 17.8; DB 17; Length 1721;  
Best Local Similarity 90.5%; Pred. No. 68;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTGTCTTAAGGAGACACCA 21  
Db 1536 GTGTCTTAAGGAGACCA 1516

## RESULT 5

US-09-997-722-4/c  
; Sequence 4, Application US/09997722  
; Publication No. US20040072154A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David  
; APPLICANT: Engelhard, Eric  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER  
; FILE REFERENCE: A-71171/RMS/DCF  
; CURRENT APPLICATION NUMBER: US/09/997,722  
; CURRENT FILING DATE: 2001-11-30  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 301  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 4  
; LENGTH: 52302  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (11301)..(11320)  
; OTHER INFORMATION: "n" at positions 11301 through 11320 can be any base.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (29267)..(29286)  
; OTHER INFORMATION: "n" at positions 29267 through 29286 can be any base.  
US-09-997-722-4

Query Match 77.4%; Score 17.8; DB 11; Length 52302;  
Best Local Similarity 90.5%; Pred. No. 87;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GTCTTAAGGAGACACACAGC 23  
Db 13323 GTTTTAAAGGAGACACAGC 13303

## RESULT 6

US-10-087-192-322

```
; Sequence 322, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, David W.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 322
; LENGTH: 144035
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-087-192-322

Query Match      77.4%; Score 17.8; DB 13; Length 144035;
Best Local Similarity 90.5%; Pred. No. 93;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 TGCTTTAAGGAGACACACAG 22
      ||||| ||||| ||||| |||||
DB      137647 TGCTTTGAAGAGACACACAG 137667

RESULT 7
US-10-741-601-9071/c
; Sequence 9071, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9071
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-601-9071

Query Match      75.7%; Score 17.4; DB 18; Length 201;
Best Local Similarity 94.7%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 TCTTAAGGAGACACACAG 22
      ||||| ||||| ||||| |||||
DB      22 TCTTCAGGAGACACACAG 4

RESULT 8
US-10-741-601-9089/c
; Sequence 9089, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9089
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-601-9089/c

Query Match      75.7%; Score 17.4; DB 18; Length 201;
Best Local Similarity 94.7%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 TCTTAAGGAGACACACAG 22
      ||||| ||||| ||||| |||||
DB      22 TCTTCAGGAGACACACAG 4

RESULT 9
US-10-741-601-9487/c
; Sequence 9487, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9487
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-601-9487

Query Match      75.7%; Score 17.4; DB 18; Length 201;
Best Local Similarity 94.7%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 TCTTAAGGAGACACACAG 22
      ||||| ||||| ||||| |||||
DB      40 TCTTCAGGAGACACACAG 22

RESULT 10
US-10-741-601-12984
; Sequence 12984, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12984
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-601-12984

Query Match      75.7%; Score 17.4; DB 18; Length 201;
Best Local Similarity 94.7%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 TCTTAAGGAGACACACAG 22
      ||||| ||||| ||||| |||||
DB      180 TCTTCAGGAGACACACAG 198

RESULT 11
US-10-741-601-12992
; Sequence 12992, Application US/10741601
```

```
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12992
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-601-12992

Query Match          75.7%; Score 17.4; DB 18; Length 201;
Best Local Similarity 94.7%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 TCTTAAGGAGACACACACAG 22
DB      130 TCTTCAGGAGACACACAG 148

RESULT 12
US-10-741-601-13191
; Sequence 13191, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13191
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-601-13191

Query Match          75.7%; Score 17.4; DB 18; Length 201;
Best Local Similarity 94.7%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 TCTTAAGGAGACACACACAG 22
DB      162 TCTTCAGGAGACACACAG 180

RESULT 13
US-10-741-600-23429/c
; Sequence 23429, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23429
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-23429

Query Match          75.7%; Score 17.4; DB 19; Length 201;
```

```
Best Local Similarity 94.7%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 TCTTAAGGAGACACACACAG 22
DB      22 TCTTCAGGAGACACACACAG 4

RESULT 14
US-10-741-600-23447/c
; Sequence 23447, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23447
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-23447

Query Match          75.7%; Score 17.4; DB 19; Length 201;
Best Local Similarity 94.7%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 TCTTAAGGAGACACACACAG 22
DB      72 TCTTCAGGAGACACACACAG 54

RESULT 15
US-10-741-600-23845/c
; Sequence 23845, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23845
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-23845

Query Match          75.7%; Score 17.4; DB 19; Length 201;
Best Local Similarity 94.7%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 TCTTAAGGAGACACACACAG 22
DB      40 TCTTCAGGAGACACACACAG 22

Search completed: March 9, 2005, 09:01:41
Job time : 510 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2005, 05:25:31 ; Search time 3237 Seconds  
(without alignments)  
270.460 Million cell updates/sec

Title: US-09-835-976B-115  
Perfect score: 23  
Sequence: 1 gtgtcttaaggagacaccacagc 23

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_gsa1:\*  
8: gb\_gsa2:\*  
9: gb\_gsa3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	19.8	86.1	911	4	BI763138
C 2	19	82.6	745	9	AG422978
C 3	18.8	81.7	811	8	BH084168
C 4	18.8	81.7	959	4	BG033867
C 5	18.8	81.7	1117	9	CL018784
C 6	18.8	81.7	1182	8	CC266894
C 7	18.4	80.0	214	4	BG100434
C 8	18.2	79.1	338	2	BE713398
C 9	18.2	79.1	631	5	BU302275
C 10	18.2	79.1	640	1	AI981983
C 11	18.2	79.1	657	8	AQ536886
C 12	18.2	79.1	698	5	BQ210941
C 13	18.2	79.1	786	7	CK454697
C 14	18.2	79.1	905	4	BG743436
C 15	18.2	79.1	1264	2	BE962595
C 16	18	78.3	360	7	CF135044
C 17	17.8	77.4	209	2	BE821011
C 18	17.8	77.4	228	7	CF920727
C 19	17.8	77.4	275	2	BE821695
C 20	17.8	77.4	323	2	AW799992
C 21	17.8	77.4	412	4	BG881935
C 22	17.8	77.4	443	2	BE684331
C 23	17.8	77.4	455	5	BQ610855
C 24	17.8	77.4	497	1	AU083746

25	17.8	77.4	501	9	CE260387	CE260387 tigr-g88-
C 26	17.8	77.4	620	5	BX257728	BX257728 BX257728
C 27	17.8	77.4	676	9	CR832914	CR832914 GROAA66B
C 28	17.8	77.4	737	8	AQ425552	AQ425552 CITBI-E1-
C 29	17.8	77.4	756	9	CE407120	CE407120 tigr-g88-
C 30	17.8	77.4	759	7	CK604020	CK604020 AGENCOURT
C 31	17.8	77.4	775	7	CN046985	CN046985 V2_p10_F1
C 32	17.8	77.4	847	7	CN047496	CN047496 V2_p13_G1
C 33	17.8	77.4	969	4	BG392149	BG392149 602410252
C 34	17.4	75.7	460	8	AZ013466	AZ013466 RPCI-23-3
C 35	17.4	75.7	534	9	CE444879	CE444879 tigr-g88-
C 36	17.4	75.7	535	4	BI969912	BI969912 GM830009B
C 37	17.4	75.7	546	8	AZ986415	AZ986415 2M0268C08
C 38	17.4	75.7	550	5	BQ273623	BQ273623 S8014E06.
C 39	17.4	75.7	555	2	AW396631	AW396631 8980609.Y
C 40	17.4	75.7	573	2	AW760528	AW760528 8151E02.Y
C 41	17.4	75.7	664	9	CR194471	CR194471 Reverse8
C 42	17.4	75.7	746	5	BQ863139	BQ863139 QSC23805.
C 43	17.4	75.7	756	9	AG600148	AG600148 Mus muscu
C 44	17.4	75.7	760	9	AG579385	AG579385 Mus muscu
C 45	17.4	75.7	761	9	AG460898	AG460898 Mus muscu

## ALIGNMENTS

RESULT 1  
BI763138/c  
LOCUS BI763138.1 GI:15754716  
DEFINITION mRNA sequence.  
ACCESSION BI763138  
VERSION BI763138.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 911)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLAM11475 row: g column: 16  
High quality sequence stop: 641.  
Location/Qualifiers  
I. .911  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5190183"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_116"  
/note="Organ: pooled colon, kidney, stomach; Vector: pCMV-SPORT6; Site: 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH\_MGC Library."

ORIGIN

```

Query Match      86.1%; Score 19.8; DB 4; Length 911;
Best Local Similarity 91.3%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTGTCTTAAGGAGACACACAGC 23
    ||||| ||||| ||||| ||||| |||||
Db 430 GTGTCTTAAGGAGACACACAGC 408

RESULT 2
AG422978      745 bp DNA linear GSS 03-JUN-2004
LOCUS      Mus musculus molossinus DNA, clone:MSMg01-294P02.T7, genomic survey
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
COMMENT

Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
BAC end Sequences of Library MSMg01
Unpublished
2 (bases 1 to 745)
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170
Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kunya Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : T7
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI.
R.Site 2 : EcoRI.
Location/Qualifiers
1. 745
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-294P02.T7"
/sex="male"
/tissue types="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

Query Match      82.6%; Score 19; DB 9; Length 745;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTCTTAAGGAGACACCA 19
    ||||| ||||| ||||| ||||| |||||
Db 69 GTGTCTTAAGGAGACACCA 87

RESULT 3
BH084168/c
LOCUS
DEFINITION
RPCI-24-366C1.TVB RPCI-24 Mus musculus genomic clone RPCI-24-366C1,
genomic survey sequence.

```

```

ACCESSION BH084168
VERSION BH084168.1 GI:14903765
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM
REFERENCE
AUTHORS
JOURNAL
COMMENT

Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other GSSs: RPCI-24-366C1.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.choi.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html
Plate: 366 row: C column: 1
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1. 811
/organism="Mus musculus"
/mol_type="genomic DNA"
/strains="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-366C1"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_lib="RPCI-24"
/notes="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

ORIGIN
Query Match      81.7%; Score 18.8; DB 8; Length 811;
Best Local Similarity 90.9%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTGTCTTAAGGAGACACACAG 22
    ||||| ||||| ||||| ||||| |||||
Db 107 GTGTCTTCGGAGACACACAG 86

RESULT 4
BG033867/c
LOCUS
DEFINITION
mRNA sequence.
ACCESSION BG033867
VERSION BG033867.1 GI:12426586
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS
JOURNAL
COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabps-r@mail.nih.gov

```

Tissue Procurement: DCTD/DTP  
 cDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM10111 row: f column: 14  
 High quality sequence stop: 756.  
 Location/Qualifiers  
 1. .959  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4402573"  
 /tissue\_type="mammary adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_87"  
 /note="Organ: breast; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.383 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."  
 ORIGIN  
 Query Match 81.7%; Score 18.8; DB 4; Length 959;  
 Best Local Similarity 90.9%; Pred. No. 3.7e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GTGCTTTAAGGAGACACACAG 22  
 Db 632 GAGTCATAAGGAGACACACAG 611  
 RESULT 5  
 CLO18784/c  
 LOCUS  
 DEFINITION  
 CH216-4B6 Sp6.1 CH216 Xenopus tropicalis genomic clone CH216-4B6,  
 genomic survey sequence.  
 ACCESSION  
 CLO18784.1 GI:40460597  
 VERSION  
 GSS.  
 KEYWORDS  
 Xenopus tropicalis (western clawed frog)  
 SOURCE  
 Xenopus tropicalis  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 Xenopodinae; Xenopus; Siurana.  
 1 (bases 1 to 1117)  
 Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,  
 Mardis, E. and Wilson, R.  
 A physical map of the xenopus tropicalis genome  
 Unpublished (2003)  
 Contact: Richard K. Wilson  
 Genome Sequencing Center  
 Washington University School of Medicine  
 Email: submissions@wustl.edu  
 Insert Length: 175000 Std Error: 0.00  
 Seq primer: Sp6 ATTTAGTGACACTATAG  
 Class: BAC ends  
 High quality sequence start: 146  
 High quality sequence stop: 428.  
 Location/Qualifiers  
 1. .1117  
 /organism="Xenopus tropicalis"  
 /mol\_type="genomic DNA"  
 /strain="Nigerian frog"  
 /db\_xref="taxon:8364"  
 /clone="CH216-4B6"  
 /sex="male"  
 /cell\_line="Stock 248 F7A2, inbred N7"  
 /clone\_lib="CH216"  
 /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis  
 BAC library"  
 FEATURES  
 source  
 1. .1117  
 /organism="Xenopus tropicalis"  
 /mol\_type="genomic DNA"  
 /strain="Nigerian frog"  
 /db\_xref="taxon:8364"  
 /clone="CH216-4B6"  
 /sex="male"  
 /cell\_line="Stock 248 F7A2, inbred N7"  
 /clone\_lib="CH216"  
 /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis  
 BAC library"

ORIGIN  
 Query Match 81.7%; Score 18.8; DB 9; Length 1117;  
 Best Local Similarity 90.9%; Pred. No. 3.8e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GTGCTTTAAGGAGACACACAG 22  
 Db 539 GTGCTTTAAGGAGACACTATAG 518  
 RESULT 6  
 CC266894  
 LOCUS  
 DEFINITION  
 CH261-154M10 Sp6.1 CH261 Gallus gallus genomic clone CH261-154M10,  
 genomic survey sequence.  
 ACCESSION  
 CC266894.1 GI:30616347  
 VERSION  
 GSS.  
 KEYWORDS  
 Gallus gallus (chicken)  
 SOURCE  
 Gallus gallus  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 1 (bases 1 to 1182)  
 Kremitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,  
 Warren, W., Graves, T., Mardis, E. and Wilson, R.  
 Gallus gallus BAC End Reads  
 Unpublished (2003)  
 Contact: Richard K. Wilson  
 Genome Sequencing Center  
 Washington University School of Medicine  
 Email: submissions@wustl.edu  
 Insert Length: 182000 Std Error: 0.00  
 Seq primer: Sp6 ATTTAGTGACACTATAG  
 Class: BAC ends  
 High quality sequence start: 207  
 High quality sequence stop: 269.  
 Location/Qualifiers  
 1. .1182  
 /organism="Gallus gallus"  
 /mol\_type="genomic DNA"  
 /strain="Red Jungle Fowl"  
 /db\_xref="taxon:9031"  
 /clone="CH261-154M10"  
 /sex="female"  
 /cell\_line="UCD001, inbred 256"  
 /clone\_lib="CH261"  
 /note="Vector: pTARBAC2.1; Site\_1: EcoRI; Site\_2: EcoRI;  
 CH261 Female Chicken library - for library and clone  
 ordering information: <http://www.chori.org/bacpac>"  
 ORIGIN  
 Query Match 81.7%; Score 18.8; DB 8; Length 1182;  
 Best Local Similarity 90.9%; Pred. No. 3.8e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GTGCTTTAAGGAGACACACAG 22  
 Db 152 GTGCTTGAAGGGGACACACAG 173  
 RESULT 7  
 BG100434/c  
 LOCUS  
 DEFINITION  
 ux85906.y1 McCarrey Eddy type B spermatogonia Mus musculus cDNA  
 clone IMAGE:3655475 5', mRNA sequence.  
 ACCESSION  
 BG100434  
 VERSION  
 BG100434.1 GI:12595751  
 KEYWORDS  
 EST.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 214)  
**REFERENCE**  
**AUTHORS**  
 Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.  
**TITLE**  
 The WashU-NCI Mouse EST Project 1999  
**JOURNAL**  
 Unpublished (1999)  
**COMMENT**  
 Contact: Marra M/WashU-NCI Mouse EST Project 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LML; contact the MGI Consortium (info@image.llnl.gov) for further information.  
 MGI:141779  
 Seq primer: Primer name ambiguous.  
**FEATURES**  
 source  
 1. .214  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="CD-1"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3655475"  
 /sex="male"  
 /tissue\_type="type B spermatogonia, pooled from multiple mice"  
 /dev\_stage="8 day"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="McCarrey Eddy type B spermatogonia"  
 /note="Organ: testis; Vector: pBluescript SK+ (Stratagene); Site 1: XhoI; Site 2: EcoRI; cDNA oligo dt-primed [5'-(GA)10-ACTAGTCGAGTGTGTTTTTTT-3'] and directionally cloned using 5' linkers 5'-AATTCGACGAG-3' and 5'-CTCGTCGC-3'. Size selection of >400bp material gives average insert size ranging from 1-2 kb. Library was mass excised (from lambda-UniZAP-XR) and resulting single-stranded phagemids were prepped and transformed into DH10B. Library contains 96% recombinants.  
 References: J. Androl. 20:635-639 and Gene 25:263-269.  
 Library constructed and donated by J. McCarrey, Ph.D. (Southwest Foundation for Biomedical Research, Dept. of Genetics); excision done by E.M. Eddy, Ph.D. (National Institutes of Health, National Institute of Environmental Health Sciences). Original lambda-based library is available through ATCC, catalog #63417."  
**ORIGIN**  
 Query Match 80.0%; Score 18.4; DB 4; Length 214;  
 Best Local Similarity 95.0%; Pred. No. 5e+02; Indels 0; Gaps 0;  
 Matches 19; Conservative 0; Mismatches 1;  
 QY 2 TGCTTTAAGGAGACACCA 21  
 |||||  
 Db 179 TGCTTTAAGGAGACACCA 160  
**RESULT 8**  
 BE713998  
**LOCUS**  
 DEFINITION  
 ME3-HT0713-080600-023-g03 HT0713 Homo sapiens cDNA, mRNA sequence.  
**ACCESSION**  
 BE713998  
**VERSION**  
 BE713998.1 GI:10102263  
**KEYWORDS**  
 EST.  
**SOURCE**  
 Homo sapiens (human)  
**ORGANISM**  
 Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE**  
 1 (bases 1 to 338)  
**AUTHORS**  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.U.  
**TITLE**  
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
**JOURNAL**  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
**MEDLINE**  
 20202663  
**PUBMED**  
 10737800  
**COMMENT**  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=MR3-HT0713-080  
 600-023-g03et3=2000-06-08et4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 7  
 High quality sequence stop: 250.  
**FEATURES**  
 source  
 1. .338  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="HT0713"  
 /note="Organ: head neck; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application NO. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
**ORIGIN**  
 Query Match 79.1%; Score 18.2; DB 2; Length 338;  
 Best Local Similarity 87.0%; Pred. No. 6.6e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 GTGCTTAAGGAGACACACAGC 23  
 |||||  
 Db 70 GTGCTTAAGGAGACACAGC 92  
**RESULT 9**  
 BU302275/c  
**LOCUS**  
 DEFINITION  
 BU302275 631 bp mRNA linear EST 27-NOV-2002  
 603741788F1 CSEQCHN56 Gallus gallus cDNA clone CHEST640022 5', mRNA sequence.  
**ACCESSION**  
 BU302275  
**VERSION**  
 BU302275.1 GI:25751924  
**KEYWORDS**  
 EST.  
**SOURCE**  
 Gallus gallus (chicken)  
**ORGANISM**  
 Gallus gallus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
**REFERENCE**  
 1 (bases 1 to 631)  
**AUTHORS**  
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
**TITLE**  
 A Comprehensive Collection of Chicken cDNAs  
**JOURNAL**  
 Curr. Biol. 12 (22), 1965-1969 (2002)  
**MEDLINE**  
 22335534  
**PUBMED**  
 12445392  
**COMMENT**  
 Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology (UMIST)



PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.

#### FEATURES

Location/Qualifiers  
1..631  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="Compton line 151"  
/db\_xref="taxon:9031"  
/clone="CHEST640o22"  
/sex="Female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="CSQCHN56"  
/note="Organ: small intestine; Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was bluntended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

#### ORIGIN

Query Match 79.1%; Score 18.2; DB 5; Length 631;  
Best Local Similarity 87.0%; Pred. No. 7e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTGTCTTAAGGAGACACACAGC 23  
||||| ||||||| |||||||  
DB 265 GTGTCTTAAGGAGACACACAGC 243

#### RESULT 10

AI981983 640 bp mRNA linear EST 07-MAY-2001  
pat.pk0070.f9.f chicken activated T cell cDNA Gallus gallus cDNA  
clone pat.pk0070.f9.f 5', mRNA sequence.

AI981983

AI981983.1 GI:5885011

EST.

Gallus gallus (chicken)

Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus.

1 (bases 1 to 640)

Tirunagaru.V.G., Sofer,L., Cui,J. and Burnside,J.

An expressed sequence tag database of T-cell-enriched activated

chicken splenocytes: sequence analysis of 5251 clones

Genomics 66 (2), 144-151 (2000)

20318616

10860659

Contact: Joan Burnside

Molecular Endocrinology

University of Delaware

40 Townsend Hall, Newark, DE 19717, USA

Tel: 302-831-1345

Fax: 302-831-3411

Email: joan@udel.edu, www.chickest.udel.edu

Seq primer: T7.

#### FEATURES

Location/Qualifiers  
1..640  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9031"

/clone="pat.pk0070.f9.f"  
/sex="male"  
/cell\_type="Con A-activated splenic T cell"  
/lab\_host="E.coli TOP10 F'"  
/clone\_lib="chicken activated T cell cDNA"  
/note="Vector: pCDNA3"

#### ORIGIN

Query Match 79.1%; Score 18.2; DB 1; Length 640;  
Best Local Similarity 87.0%; Pred. No. 7.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTGTCTTAAGGAGACACACAGC 23  
||||| ||||||| |||||||  
DB 403 GTGTCTTAAGGAGACTGCACAGC 425

#### RESULT 11

AQ536886

LOCUS

DEFINITION

657 bp DNA linear GSS 18-MAY-1999

RPCI-11-319E4.TV RPCI-11 Homo sapiens genomic clone RPCI-11-319E4,

genomic survey sequence.

ACCESSION

AQ536886

VERSION

AQ536886.1 GI:4848576

KEYWORDS

GSS.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 657)

Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and

Venter,J.C.

Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

Map Building

Unpublished (1997)

Other\_GSSs: RPCI-11-319E4.TJ

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hb@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from

Research Genet cs (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/hungen/bac\_end\_search/bac\_end\_search.html.

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1..657

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="GDB:7622211"

/db\_xref="taxon:9606"

/clone="RPCI-11-319E4"

/sex="Male"

/cell\_type="Lymphocytes"

/clone\_lib="RPCI-11"

/note="Vector: pBACe3.6; Site1: EcoRI; Site2: EcoRI;

RPCI11 Human Male BAC Library"

#### ORIGIN

Query Match 79.1%; Score 18.2; DB 8; Length 657;  
Best Local Similarity 87.0%; Pred. No. 7.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTGTCTTAAGGAGACACACAGC 23  
||||| ||||||| |||||||  
DB 424 GTGTCTTAAGAATCACCACAGC 446

RESULT 12  
BQ210941/c  
LOCUS  
DEFINITION  
UT-R-DY1-coj-e-09-0-UI.sl NCI\_CGAP\_DY1 Rattus norvegicus cDNA clone  
IMAGE:7336019.3', mRNA sequence.

ACCESSION  
BQ210941  
VERSION  
BQ210941.1 GI:20427406  
KEYWORDS  
EST.  
SOURCE  
Rattus norvegicus (Norway rat)  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
1 (bases 1 to 698)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
AUTHORS  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
TITLE  
Normalizing and subtraction: two approaches to facilitate gene  
JOURNAL  
Genome Res. 6 (9), 791-806 (1996)  
MEDLINE  
97044477  
PUBMED  
8889548  
COMMENT  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics / NIH  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: cgabs-r@mail.nih.gov  
Oligo-dT track not found, Not I site shown in beginning of sequence  
is likely internal to the message. cDNA library Preparation: M.B.  
Soares Lab Clone distribution: clones will be available through  
IMAGE (http://image.llnl.gov)  
Seq primer: M13 Forward  
POLYA=Yes.

FEATURES  
source

1..698  
Location/Qualifiers  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="IMAGE:7336019"  
/issue\_type="Cartilage"  
/dev\_stage="ADULT"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI\_CGAP\_DY1"  
/notes="Organ: Femur and Tibia; Vector: pTT3D-Pac  
(Pharmacia) with a modified polylinker; Site 1: Not I;  
Site 2: Eco RI; UI-R-DY1 is a normalized cDNA library  
containing the following tissue(s): Rat Cartilage from  
Femur and Tibia. The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was ligated to an EcoR I adaptor, digested with Not  
I, and cloned directionally into pTT3-Pac vector. The  
oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dT)18 tail. The  
sequence tag for this library is CTAATGGACG. The Rat  
cartilage tissue was provided by Dr Jeff Stevens at the  
University of Iowa.  
TAG\_TISSUE=cartilage  
TAG\_LIB=UI-R-DY1  
TAG\_SEQ=CTAATGGACG"

## ORIGIN

Query Match 79.1%; Score 18.2; DB 5; Length 698;  
Best Local Similarity 87.0%; Pred. No. 7.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTGTCTTAAGGAGACACACG 23  
|||||  
Db 697 GTGTCTTAAGGAGACACACG 675

RESULT 13  
CK454697  
LOCUS  
DEFINITION  
915199 MARC 4PIG Sus scrofa cDNA 5', mRNA sequence.  
ACCESSION  
CK454697  
VERSION  
CK454697.1 GI:40801911  
KEYWORDS  
EST.  
SOURCE  
Sus scrofa (pig)  
ORGANISM  
Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
REFERENCE  
1 (bases 1 to 786)  
Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,  
Nonneman,D.J., Wray,J.E. and Keeler,J.W.  
AUTHORS  
Porcine EST collection using a normalized library constructed from  
embryos representing early developmental stages  
TITLE  
Unpublished (2003)  
JOURNAL  
Contact: Smith TPL  
COMMENT  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called with phred v0.020425.c and  
trimmed with the aid of the trim\_alt option. Vector identified with  
cross match v0.990329.  
Plate: TMM8026 row: 0 column: 15  
Seq primer: GTAATACGACTCCTACTATAGG.

FEATURES  
source

1..786  
Location/Qualifiers  
/organism="Sus scrofa"  
/mol\_type="mRNA"  
/db\_xref="taxon:9823"  
/issue\_type="pooled"  
/lab\_host="DH10B"  
/clone\_lib="MARC 4PIG"  
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;  
Library made with combined RNA from day-10, day-13,  
day-15, day-25, and day-30 whole embryos."

## ORIGIN

Query Match 79.1%; Score 18.2; DB 7; Length 786;  
Best Local Similarity 87.0%; Pred. No. 7.2e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTGTCTTAAGGAGACACACG 23  
|||||  
Db 270 GTGTCTTAAGGAGACACACG 292

RESULT 14  
BG743436/c

LOCUS  
DEFINITION  
602634952F1 NCI\_CGAP\_Skn3 Homo sapiens cDNA clone IMAGE:4780069 5',  
mRNA sequence.

ACCESSION  
BG743436  
VERSION  
BG743436.1 GI:14054089  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 905)  
NIH-MGC http://mgc.nci.nih.gov/  
AUTHORS  
National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE  
Unpublished (1999)  
JOURNAL  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: James Cleaver, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA  
Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10638 row: 9 column: 14  
High quality sequence stop: 757.

## FEATURES

## source

1. .905  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4780069"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI CGAP SKn3"  
/note="Organ: skin; Vector: pCMV-SPORT6; Site: 1: NotI;  
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dr.  
Average insert size 1.5kb. Library constructed by Life  
Technologies. Note: this is a NCI CGAP Library."

## ORIGIN

Query Match 79.1%; Score 18.2; DB 4; Length 905;  
Best Local Similarity 87.0%; Pred. No. 7.3e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTGTCTTAAGGAGACACACAGC 23

Db 738 GTGTCTTAAGGAGACACACAGC 716

## RESULT 15

BE962595  
LOCUS 1264 bp mRNA linear EST 14-DEC-2000  
DEFINITION 601855938R1 NIH\_MGC\_66 Homo sapiens cDNA clone IMAGE:3855873 3',  
mRNA sequence.  
ACCESSION BE962595  
VERSION BE962595.2 GI:11765651  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1264)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
On Oct 3, 2000 this sequence version replaced gi:10573300.  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov  
Tissue Procurement: DCTD/DTF  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM570 row: k column: 10  
High quality sequence stop: 77.

## FEATURES

## source

1. .1264  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3855873"  
/tissue type="adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 66"  
/note="Organ: ovary; Vector: pCMV-SPORT6; Site: 1: NotI;  
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dr.  
Average insert size 1.8 kb. Library constructed by Life  
Technologies."

## ORIGIN

Query Match 79.1%; Score 18.2; DB 2; Length 1264;  
Best Local Similarity 87.0%; Pred. No. 7.6e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTGTCTTAAGGAGACACACAGC 23  
Db 142 GTGTCTTAAGGAGACACACAGC 164  
Search completed: March 9, 2005, 08:47:38  
Job time : 3241 secs

This Page Blank (uspto)

This Page Blank (uspto)

This Page Blank (uspto)